(19) World Intellectual Property Organization International Bureau



(43) International Publication Date 21 December 2000 (21.12.2000)

PCT

(10) International Publication Number WO 00/77239 A2

(51) International Patent Classification7:

C12Q

- (21) International Application Number: PCT/US00/14858
- (22) International Filing Date: 24 May 2000 (24.05.2000)
- (25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data: 09/333,159

14 June 1999 (14.06.1999) US

- (71) Applicant: MILLENNIUM PHARMACEUTICALS, INC. [US/US]; 75 Sidney Street, Cambridge, MA 02139 (US).
- (72) Inventors: MCCARTHY, Sean, A.; 62 Commonwealth Avenue #4, Boston, MA 02116 (US). FRASER, Christopher, C.; 52 Grassland Street, Lexington, MA 02421 (US). SHARP, John, D.; 245 Park Avenue, Arlington, MA 02476 (US). BARNES, Thomas, M.; 22 Hanson Street #2, Boston, MA 02118 (US).

- (74) Agents: COLBY, Gary, D. et al.; Akin, Gump, Strauss, Hauer & Feld, L.L.P., One Commerce Square, Suite 2200, 2005 Market Street, Philadelphia, PA 19103-7086 (US).
- (81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW.
- (84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:

 Without international search report and to be republished upon receipt of that report.

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: NOVEL GENES ENCODING PROTEINS HAVING DIAGNOSTIC, PREVENTIVE, THERAPEUTIC, AND OTHER USES

(57) Abstract: The invention provides isolated nucleic acids encoding a variety of proteins having diagnostic, preventive, therapeutic, and other uses. These nucleic and proteins are useful for diagnosis, prevention, and therapy of a number of human and other animal disorders. The invention also provides antisense nucleic acid molecules, expression vectors containing the nucleic acid molecules of the invention, host cells into which the expression vectors have been introduced, and non-human transgenic animals in which a nucleic acid molecule of the invention has been introduced or disrupted. The invention still further provides isolated polypeptides, fusion polypeptides, antigenic peptides and antibodies. Diagnostic, screening, and therapeutic methods utilizing compositions of the invention are also provided. The nucleic acids and polypeptides of the present invention are useful as modulating agents in regulating a variety of cellular processes.

19 A2

NOVEL GENES ENCODING PROTEINS HAVING DIAGNOSTIC, PREVENTIVE, THERAPEUTIC, AND OTHER USES

Cross Reference to Related Applications

This application is a continuation-in-part of co-pending United

10 States Patent application number 09/333,159, filed June 14, 1999.

5

15

. 20

25

30

Background of the Invention

The molecular bases underlying many human and animal physiological states (e.g., diseased and homeostatic states of various tissues) remain unknown. Nonetheless, it is well understood that these states result from interactions among the proteins and nucleic acids present in the cells of the relevant tissues. In the past, the complexity of biological systems overwhelmed the ability of practitioners to understand the molecular interactions giving rise to normal and abnormal physiological states. More recently, though, the techniques of molecular biology, transgenic and null mutant animal production, computational biology, pharmacogenomics, and the like have enabled practitioners to discern the role and importance of individual genes and proteins in particular physiological states.

Knowledge of the sequences and other properties of genes (particularly including the portions of genes encoding proteins) and the proteins encoded thereby enables the practitioner to design and screen agents which will affect, prospectively or retrospectively, the physiological state of an animal tissue in a favorable way. Such knowledge also enables the practitioner, by detecting the levels of gene expression and protein production, to diagnose the current physiological state of a tissue or animal and to predict such physiological states in the future. This knowledge furthermore enables the practitioner to identify and design molecules which bind with the polynucleotides and proteins, *in vitro*, *in vivo*, or both.

The present invention provides sequence information for polynucleotides derived from human and murine genes and for proteins encoded

thereby, and thus enables the practitioner to assess, predict, and affect the physiological state of various human and murine tissues.

5

10

15

20

25

30

Summary of the Invention

The present invention is based, at least in part, on the discovery of a variety of human and murine cDNA molecules which encode proteins which are herein designated TANGO 202, TANGO 234, TANGO 265, TANGO 273, TANGO 286, TANGO 294, and INTERCEPT 296. These seven proteins, fragments thereof, derivatives thereof, and variants thereof are collectively referred to herein as the polypeptides of the invention or the proteins of the invention. Nucleic acid molecules encoding polypeptides of the invention are collectively referred to as nucleic acids of the invention.

The nucleic acids and polypeptides of the present invention are useful as modulating agents in regulating a variety of cellular processes.

Accordingly, in one aspect, the present invention provides isolated nucleic acid molecules encoding a polypeptide of the invention or a biologically active portion thereof. The present invention also provides nucleic acid molecules which are suitable as primers or hybridization probes for the detection of nucleic acids encoding a polypeptide of the invention.

The invention also features nucleic acid molecules which are at least 40% (or 50%, 60%, 70%, 80%, 90%, 95%, or 98%) identical to the nucleotide sequence of any of SEQ ID NOs: 1, 2, 9, 10, 17, 18, 25, 26, 33, 34, 45, 46, 53, 54, 67, 68, 72, and 73, the nucleotide sequence of a cDNA clone deposited with ATCC® as one of Accession numbers 207219, 207184, 207228, 207185, 207220, and 207221 ("a cDNA of a clone deposited as one of ATCC® 207219, 207184, 207228, 207185, 207220, and 207221"), or a complement thereof.

The invention features nucleic acid molecules which include a fragment of at least 15 (25, 40, 60, 80, 100, 150, 200, 250, 300, 350, 400, 450, 550, 650, 700, 800, 900, 1000, 1200, 1400, 1600, 1800, 2000, 2200, 2400, 2600, 2800, 3000, 3500, 4000, 4500, or 4928) consecutive nucleotide residues of any of SEQ ID NOs: 1, 2, 9, 10, 17, 18, 25, 26, 33, 34, 45, 46, 53, 54, 67, 68, 72, and 73, the

nucleotide sequence of a cDNA of a clone deposited as one of ATCC® 207219, 207184, 207228, 207185, 207220, and 207221, or a complement thereof.

The invention also features nucleic acid molecules which include a nucleotide sequence encoding a protein having an amino acid sequence that is at least 50% (or 60%, 70%, 80%, 90%, 95%, or 98%) identical to the amino acid sequence of any of SEQ ID NOs: 3-8, 11-16, 19-24, 27-32, 35-44, 47-52, 55-66, 69, and 74, or the amino acid sequence encoded by a cDNA of a clone deposited as one of ATCC® 207219, 207184, 207228, 207185, 207220, and 207221, or a complement thereof.

5

10

15

20

25

30

In preferred embodiments, the nucleic acid molecules have the nucleotide sequence of any of SEQ ID NOs: 1, 2, 9, 10, 17, 18, 25, 26, 33, 34, 45, 46, 53, 54, 67, 68, 72, and 73, or the nucleotide sequence of a cDNA of a clone deposited as one of ATCC[®] 207219, 207184, 207228, 207185, 207220, and 207221.

Also within the invention are nucleic acid molecules which encode a fragment of a polypeptide having the amino acid sequence of any of SEQ ID NOs: 3-8, 11-16, 19-24, 27-32, 35-44, 47-52, 55-66, 69, and 74, or the amino acid sequence encoded by a cDNA of a clone deposited as one of ATCC® 207219, 207184, 207228, 207185, 207220, and 207221, the fragment including at least 8 (10, 15, 20, 25, 30, 40, 50, 75, 100, 125, 150, or 200) consecutive amino acids of any of SEQ ID NOs: 3-8, 11-16, 19-24, 27-32, 35-44, 47-52, 55-66, 69, and 74, or the amino acid sequence encoded by a cDNA of a clone deposited as one of ATCC® 207219, 207184, 207228, 207185, 207220, and 207221.

The invention includes nucleic acid molecules which encode a naturally occurring allelic variant of a polypeptide comprising the amino acid sequence of any of SEQ ID NOs: 3-8, 11-16, 19-24, 27-32, 35-44, 47-52, 55-66, 69, and 74, or the amino acid sequence encoded by a cDNA of a clone deposited as one of ATCC® 207219, 207184, 207228, 207185, 207220, and 207221, wherein the nucleic acid molecule hybridizes under stringent conditions to a nucleic acid molecule having a nucleic acid sequence encoding any of SEQ ID NOs: 1, 2, 9, 10, 17, 18, 25, 26, 33, 34, 45, 46, 53, 54, 67, 68, 72, and 73, the nucleotide sequence of

a cDNA of a clone deposited as one of ATCC® 207219, 207184, 207228, 207185, 207220, and 207221, or a complement thereof.

Also within the invention are isolated polypeptides or proteins having an amino acid sequence that is at least about 50%, preferably 60%, 75%, 90%, 95%, or 98% identical to the amino acid sequence of any of SEQ ID NOs: 3-8, 11-16, 19-24, 27-32, 35-44, 47-52, 55-66, 69, and 74.

5

10

15

20

25

30

Also within the invention are isolated polypeptides or proteins which are encoded by a nucleic acid molecule having a nucleotide sequence that is at least about 40%, preferably 50%, 75%, 85%, or 95% identical the nucleic acid sequence encoding any of SEQ ID NOs: 3-8, 11-16, 19-24, 27-32, 35-44, 47-52, 55-66, 69, and 74, and isolated polypeptides or proteins which are encoded by a nucleic acid molecule consisting of the nucleotide sequence which hybridizes under stringent hybridization conditions to a nucleic acid molecule having the nucleotide sequence of any of SEQ ID NOs: 1, 2, 9, 10, 17, 18, 25, 26, 33, 34, 45, 46, 53, 54, 67, 68, 72, and 73.

Also within the invention are polypeptides which are naturally occurring allelic variants of a polypeptide that includes the amino acid sequence of any of SEQ ID NOs: 3-8, 11-16, 19-24, 27-32, 35-44, 47-52, 55-66, 69, and 74, or the amino acid sequence encoded by a cDNA of a clone deposited as one of ATCC[®] 207219, 207184, 207228, 207185, 207220, and 207221, wherein the polypeptide is encoded by a nucleic acid molecule which hybridizes under stringent conditions to a nucleic acid molecule having the nucleotide sequence of any of SEQ ID NOs: 1, 2, 9, 10, 17, 18, 25, 26, 33, 34, 45, 46, 53, 54, 67, 68, 72, and 73, or a complement thereof.

The invention also features nucleic acid molecules that hybridize under stringent conditions to a nucleic acid molecule having the nucleotide sequence of any of SEQ ID NOs: 1, 2, 9, 10, 17, 18, 25, 26, 33, 34, 45, 46, 53, 54, 67, 68, 72, and 73, the nucleotide sequence of a cDNA of a clone deposited as one of ATCC® 207219, 207184, 207228, 207185, 207220, and 207221, or a complement thereof. In other embodiments, the nucleic acid molecules are at least 15 (25, 40, 60, 80, 100, 150, 200, 250, 300, 350, 400, 450, 550, 650, 700, 800, 900, 1000, 1200, 1400, 1600, 1800, 2000, 2200, 2400, 2600, 2800, 3000, 3500, 4000,

4500, or 4928) nucleotides in length and hybridize under stringent conditions to a nucleic acid molecule having the nucleotide sequence of any of SEQ ID NOs: 1, 2, 9, 10, 17, 18, 25, 26, 33, 34, 45, 46, 53, 54, 67, 68, 72, and 73, the nucleotide sequence of a cDNA of a clone deposited as one of ATCC[®] 207219, 207184, 207228, 207185, 207220, and 207221, or a complement thereof. In some embodiments, the isolated nucleic acid molecules encode a cytoplasmic, transmembrane, extracellular, or other domain of a polypeptide of the invention. In other embodiments, the invention provides an isolated nucleic acid molecule which is antisense to the coding strand of a nucleic acid of the invention.

10

15

20

5

Another aspect of the invention provides vectors, e.g., recombinant expression vectors, comprising a nucleic acid molecule of the invention. In another embodiment, the invention provides isolated host cells, e.g., mammalian and non-mammalian cells, containing such a vector or a nucleic acid of the invention. The invention also provides methods for producing a polypeptide of the invention by culturing, in a suitable medium, a host cell of the invention containing a recombinant expression vector encoding a polypeptide of the invention such that the polypeptide of the invention is produced.

Another aspect of this invention features isolated or recombinant proteins and polypeptides of the invention. Preferred proteins and polypeptides possess at least one biological activity possessed by the corresponding naturally-occurring human polypeptide. An activity, a biological activity, and a functional activity of a polypeptide of the invention refers to an activity exerted by a protein or polypeptide of the invention on a responsive cell as determined *in vivo*, or *in vitro*, according to standard techniques.

25

30

Such activities can be a direct activity, such as an association with or an enzymatic activity on a second protein, or an indirect activity, such as a cellular process (e.g., signaling activity) mediated by interaction of the protein with a second protein. Such activities include, by way of example, formation of protein-protein interactions with proteins of one or more signaling pathways (e.g., with a protein with which the naturally-occurring polypeptide interacts); binding with a ligand of the naturally-occurring protein; and binding with an intracellular target of the naturally-occurring protein. Other activities include modulation of one or more

of cellular proliferation, of cellular differentiation, of chemotaxis, of cellular migration, and of cell death (e.g., apoptosis).

5

10

15

20

25

30

By way of example, TANGO 202 exhibits the ability to affect growth, proliferation, survival, differentiation, and activity of human hematopoietic cells (e.g., bone marrow stromal cells) and fetal cells. TANGO 202 modulates cellular binding to one or more mediators, modulates proteolytic activity *in vivo*, modulates developmental processes, and modulates cell growth, proliferation, survival, differentiation, and activity. Thus, TANGO 202 can be used to prevent, diagnose, or treat disorders relating to aberrant cellular protease activity, inappropriate interaction (or non-interaction) of cells with mediators, inappropriate development, and blood and hematopoietic cell-related disorders. Exemplary disorders for which TANGO 202 is useful include immune disorders, infectious diseases, auto-immune disorders, vascular and cardiovascular disorders, disorders related to mal-expression of growth factors, cancers, hematological disorders, various cancers, birth defects, developmental defects, and the like.

Further by way of example, TANGO 234 exhibits the ability to affect growth, proliferation, survival, differentiation, and activity of human lung, hematopoietic, and fetal cells and of (e.g., bacterial or fungal) cells and viruses which infect humans. TANGO 234 modulates growth, proliferation, survival, differentiation, and activity of gamma delta T cells, for example. Furthermore, TANGO 234 modulates cholesterol deposition on human arterial walls, and is involved in uptake and metabolism of low density lipoprotein and regulation of serum cholesterol levels.

Thus, TANGO 234 can be used to affect development and persistence of atherogenesis and arteriosclerosis, as well as other vascular and cardiovascular disorders. Other exemplary disorders for which TANGO 234 is useful include immune development disorders and disorders involving generation and persistence of an immune response to bacterial, fungal, and viral infections.

Still further by way of example, TANGO 265 modulates growth and regeneration of neuronal and epithelial tissues, and guides neuronal axon development. TANGO 265 is a transmembrane protein which mediates cellular interaction with cells, molecules and structures (e.g., extracellular matrix) in the

extracellular environment. TANGO 265 is therefore involved in growth, organization, and adhesion of tissues and the cells which constitute those tissues. Furthermore, TANGO 265 modulates growth, proliferation, survival, differentiation, and activity of neuronal cells and immune system cells. Thus, TANGO 265 can be used, for example, to prevent, diagnose, or treat disorders characterized by aberrant organization or development of a tissue or organ, for guiding neural axon development, for modulating differentiation of cells of the immune system, for modulating cytokine production by cells of the immune system, for modulating reactivity of cells of the immune system toward cytokines, for modulating initiation and persistence of an inflammatory response, and for modulating proliferation of epithelial cells.

5

10

15

20

25

30

Yet further by way of example, TANGO 273 protein mediates one or more physiological responses of cells to bacterial infection, e.g., by mediating one or more of detection of bacteria in a tissue in which it is expressed, movement of cells with relation to sites of bacterial infection, production of biological molecules which inhibit bacterial infection, and production of biological molecules which alleviate cellular or other physiological damage wrought by bacterial infection. TANGO 273, a transmembrane protein, is also involved in transmembrane signal transduction, and therefore mediates transmission of signals between the extracellular and intracellular environments of cells. TANGO 273 mediates regulation of cell growth and proliferation, endocytosis, activation of respiratory burst, and other physiological processes triggered by transmission of a signal via a protein with which TANGO 273 interacts. The compositions and methods of the invention can therefore be used to prevent, diagnose, and treat disorders involving one or more physiological activities mediated by TANGO 273 protein. Such disorders include, for example, various bone-related disorders such as metabolic, homeostatic, and developmental bone disorders (e.g., osteoporosis, various cancers, skeletal development disorders, bone fragility and the like), disorders caused by or related to bacterial infection, and disorders characterized by aberrant transmembrane signal transduction by TANGO 273.

As an additional example, TANGO 286 protein is involved in lipidbinding physiological processes such as lipid transport, metabolism, serum lipid

particle regulation, host anti-microbial defensive mechanisms, and the like. Thus, the compositions and methods of the invention can therefore be used to prevent, diagnose, and treat disorders involving one or more physiological activities mediated by TANGO 286 protein. Such disorders include, for example, lipid transport disorders, lipid metabolism disorders, obesity, disorders of serum lipid particle regulation, disorders involving insufficient or inappropriate host anti-microbial defensive mechanisms, vasculitis, bronchiectasis, LPS-related disorders such as shock, disseminated intravascular coagulation, anemia, thrombocytopenia, adult respiratory distress syndrome, renal failure, liver disease, and disorders associated with Gram negative bacterial infections, such as bacteremia, endotoxemia, sepsis, and the like.

5

10

15

20

25

30

Further by way of example, TANGO 294 protein is involved in facilitating absorption and metabolism of fat. Thus, the compositions and methods of the invention can therefore be used to prevent, diagnose, and treat disorders involving one or more physiological activities mediated by TANGO 294 protein. Such disorders include, for example, inadequate expression of gastric/pancreatic lipase, cystic fibrosis, exocrine pancreatic insufficiency, medical treatments which alter fat absorption, obesity, and the like.

As another example, INTERCEPT 296 protein is involved in physiological processes related to disorders of the human lung and esophagus. Thus, the compositions and methods of the invention can be used to prevent, diagnose, and treat these disorders. Such disorders include, for example, various cancers, bronchitis, cystic fibrosis, respiratory infections (e.g., influenza, bronchiolitis, pneumonia, and tuberculosis), asthma, emphysema, chronic bronchitis, bronchiectasis, pulmonary edema, pleural effusion, pulmonary embolus, adult and infant respiratory distress syndromes, heartburn, and gastric reflux esophageal disease.

In one embodiment, a polypeptide of the invention has an amino acid sequence sufficiently identical to an identified domain of a polypeptide of the invention. As used herein, the term "sufficiently identical" refers to a first amino acid or nucleotide sequence which contains a sufficient or minimum number of identical or equivalent (e.g., with a similar side chain) amino acid residues or

nucleotides to a second amino acid or nucleotide sequence such that the first and second amino acid or nucleotide sequences have a common structural domain and/or common functional activity. For example, amino acid or nucleotide sequences which contain a common structural domain having about 65% identity, preferably 75% identity, more preferably 85%, 95%, or 98% identity are defined herein as sufficiently identical.

5

10

15

20

25

30

In one embodiment, the isolated polypeptide of the invention lacks both a transmembrane and a cytoplasmic domain. In another embodiment, the polypeptide lacks both a transmembrane domain and a cytoplasmic domain and is soluble under physiological conditions.

The polypeptides of the present invention, or biologically active portions thereof, can be operably linked to a heterologous amino acid sequence to form fusion proteins. The invention further features antibody substances that specifically bind a polypeptide of the invention such as monoclonal or polyclonal antibodies, antibody fragments, single-chain antibodies, and the like. In addition, the polypeptides of the invention or biologically active portions thereof can be incorporated into pharmaceutical compositions, which optionally include pharmaceutically acceptable carriers. These antibody substances can be made, for example, by providing the polypeptide of the invention to an immunocompetent vertebrate and thereafter harvesting blood or serum from the vertebrate.

In another aspect, the present invention provides methods for detecting the presence of the activity or expression of a polypeptide of the invention in a biological sample by contacting the biological sample with an agent capable of detecting an indicator of activity such that the presence of activity is detected in the biological sample.

In another aspect, the invention provides methods for modulating activity of a polypeptide of the invention comprising contacting a cell with an agent that modulates (inhibits or enhances) the activity or expression of a polypeptide of the invention such that activity or expression in the cell is modulated. In one embodiment, the agent is an antibody that specifically binds to a polypeptide of the invention.

In another embodiment, the agent modulates expression of a polypeptide of the invention by modulating transcription, splicing, or translation of an mRNA encoding a polypeptide of the invention. In yet another embodiment, the agent is a nucleic acid molecule having a nucleotide sequence that is antisense with respect to the coding strand of an mRNA encoding a polypeptide of the invention.

5

10

15

20

25

30

The present invention also provides methods to treat a subject having a disorder characterized by aberrant activity of a polypeptide of the invention or aberrant expression of a nucleic acid of the invention by administering an agent which is a modulator of the activity of a polypeptide of the invention or a modulator of the expression of a nucleic acid of the invention to the subject. In one embodiment, the modulator is a protein of the invention. In another embodiment, the modulator is a nucleic acid of the invention. In other embodiments, the modulator is a peptide, peptidomimetic, or other small molecule (e.g., a small organic molecule).

The present invention also provides diagnostic assays for identifying the presence or absence of a genetic lesion or mutation characterized by at least one of: (i) aberrant modification or mutation of a gene encoding a polypeptide of the invention, (ii) mis-regulation of a gene encoding a polypeptide of the invention, and (iii) aberrant post-translational modification of a polypeptide of the invention wherein a wild-type form of the gene encodes a polypeptide having the activity of the polypeptide of the invention.

In another aspect, the invention provides a method for identifying a compound that binds to or modulates the activity of a polypeptide of the invention. In general, such methods entail measuring a biological activity of the polypeptide in the presence and absence of a test compound and identifying those compounds which alter the activity of the polypeptide.

The invention also features methods for identifying a compound which modulates the expression of a polypeptide or nucleic acid of the invention by measuring the expression of the polypeptide or nucleic acid in the presence and absence of the compound.

In yet a further aspect, the invention provides substantially purified antibodies or fragments thereof (i.e., antibody substances), including non-human

antibodies or fragments thereof, which specifically bind with a polypeptide of the invention or with a portion thereof. In various embodiments, these substantially purified antibodies/fragments can be human, non-human, chimeric, and/or humanized antibodies. Non-human antibodies included in the invention include, by way of example, goat, mouse, sheep, horse, chicken, rabbit, and rat antibodies. In addition, the antibodies of the invention can be polyclonal antibodies or monoclonal antibodies.

5

10

15

20

25

30

In a particularly preferred embodiment, the antibody substance of the invention specifically binds with an extracellular domain of one of TANGO 202, TANGO 234, TANGO 265, TANGO 273, TANGO 286, TANGO 294, and INTERCEPT 296. Preferably, the extracellular domain with which the antibody substance binds has an amino acid sequence selected from the group consisting of SEQ ID NOs: 5, 6, 14, 22, 30, 37, 49, 50, and 56-58.

Any of the antibody substances of the invention can be conjugated with a therapeutic moiety or with a detectable substance. Non-limiting examples of detectable substances that can be conjugated with the antibody substances of the invention include an enzyme, a prosthetic group, a fluorescent material (i.e., a fluorophore), a luminescent material, a bioluminescent material, and a radioactive material (e.g., a radionuclide or a substituent comprising a radionuclide).

The invention also provides a kit containing an antibody substance of the invention conjugated with a detectable substance, and instructions for use. Still another aspect of the invention is a pharmaceutical composition comprising an antibody substance of the invention and a pharmaceutically acceptable carrier. In preferred embodiments, the pharmaceutical composition contains an antibody substance of the invention, a therapeutic moiety (preferably conjugated with the antibody substance), and a pharmaceutically acceptable carrier.

Other features and advantages of the invention will be apparent from the following detailed description and claims.

Brief Description of the Drawings

Figure 1 comprises Figures 1A-1M. The nucleotide sequence (SEQ ID NO: 1) of a cDNA encoding the human TANGO 202 protein described herein is

5

10

15

20

25

30

listed in Figures 1A-1D. The open reading frame (ORF; residues 34 to 1458; SEQ ID NO: 2) of the cDNA is indicated by nucleotide triplets, above which the amino acid sequence (SEQ ID NO: 3) of human TANGO 202 is listed. The nucleotide sequence (SEQ ID NO: 67) of a cDNA encoding the murine TANGO 202 protein described herein is listed in Figures 1E-1I. The ORF (residues 81 to 1490; SEQ ID NO: 68) of the cDNA is indicated by nucleotide triplets, above which the amino acid sequence (SEQ ID NO: 69) of murine TANGO 202 is listed. An alignment of the amino acid sequences of human ("Hum."; SEQ ID NO: 3) and murine ("Mur."; SEO ID NO: 69) TANGO 202 protein is shown in Figures 1J and 1K, wherein identical amino acid residues are indicated by ":" and similar amino acid residues are indicated by ".". Figure 1L is a hydrophilicity plot of human TANGO 202 protein, in which the locations of cysteine residues ("Cys") and potential Nglycosylation sites ("Ngly") are indicated by vertical bars and the predicted extracellular ("out"), intracellular ("ins"), or transmembrane ("TM") locations of the protein backbone is indicated by a horizontal bar. Figure 1M is a hydrophilicity plot of murine TANGO 202 protein.

Figure 2 comprises Figures 2A-2Qxvii. The nucleotide sequence (SEQ ID NO: 9) of a cDNA encoding the human TANGO 234 protein described herein is listed in Figures 2A-2I. The ORF (residues 28 to 4386; SEQ ID NO: 10) of the cDNA is indicated by nucleotide triplets, above which the amino acid sequence (SEQ ID NO: 11) of human TANGO 234 is listed. Figure 2J is a hydrophilicity plot of human TANGO 234 protein. An alignment of the amino acid sequences of human TANGO 234 ("Hum"; SEQ ID NO: 11) and bovine WC1 ("WC1"; SEQ ID NO: 78) proteins is shown in Figures 2K-2P, wherein identical amino acid residues are indicated by ":" and similar amino acid residues are indicated by ".". An alignment of the nucleotide sequences of an ORF encoding human TANGO 234 ("Hum"; SEQ ID NO: 10) and an ORF encoding bovine WC1 ("WC1"; SEQ ID NO: 79) proteins is shown in Figures 2Qi-2Qxvii, wherein identical nucleotide residues are indicated by ":".

Figure 3 comprises Figures 3A-3U. The nucleotide sequence (SEQ ID NO: 17) of a cDNA encoding the human TANGO 265 protein described herein is listed in Figures 3A-3E. The ORF (residues 32 to 2314; SEQ ID NO: 18) of the

cDNA is indicated by nucleotide triplets, above which the amino acid sequence (SEQ ID NO: 19) of human TANGO 265 is listed. An alignment of the amino acid sequences of human TANGO 265 protein ("Hum."; SEQ ID NO: 19) and murine semaphorin B protein ("Mur."; SEQ ID NO: 70; GenBank Accession No. X85991) is shown in Figures 3F-3H, wherein identical amino acid residues are indicated by ":" and similar amino acid residues are indicated by ".". In Figures 3I-3T, an alignment of the nucleotide sequences of the cDNA encoding human TANGO 265 protein ("Hum."; SEQ ID NO: 17) and the nucleotide sequences of the cDNA encoding murine semaphorin B protein ("Mur."; SEQ ID NO: 71; GenBank Accession No. X85991) is shown. Figure 3U is a hydrophilicity plot of TANGO 265 protein.

5

10

15

20

25

30

ID NO: 25) of a cDNA encoding the human TANGO 273 protein described herein is listed in Figures 4A-4C. The ORF (residues 135 to 650; SEQ ID NO: 26) of the cDNA is indicated by nucleotide triplets, above which the amino acid sequence (SEQ ID NO: 27) of human TANGO 273 is listed. The nucleotide sequence (SEQ ID NO: 72) of a cDNA encoding the murine TANGO 273 protein described herein is listed in Figures 4D-4G. The ORF (residues 137 to 652; SEQ ID NO: 73) of the cDNA is indicated by nucleotide triplets, above which the amino acid sequence (SEQ ID NO: 74) of murine TANGO 273 is listed. An alignment of the amino acid sequences of human ("Hum."; SEQ ID NO: 27) and murine ("Mur."; SEQ ID NO: 74) TANGO 273 protein is shown in Figure 4H, wherein identical amino acid residues are indicated by ":" and similar amino acid residues are indicated by ":" and similar amino acid residues are indicated by ".".

Figure 4I is a hydrophilicity plot of human TANGO 273 protein, and Figure 4J is a hydrophilicity plot of murine TANGO 273 protein.

Figure 5 comprises Figures 5A-5I. The nucleotide sequence (SEQ ID NO: 33) of a cDNA encoding the human TANGO 286 protein described herein is listed in Figures 5A-5D. The ORF (residues 133 to 1497; SEQ ID NO: 34) of the cDNA is indicated by nucleotide triplets, above which the amino acid sequence (SEQ ID NO: 35) of human TANGO 286 is listed. Figure 5E is a hydrophilicity plot of TANGO 286 protein. An alignment of the amino acid sequences of human TANGO 286 ("286"; SEQ ID NO: 35) and BPI protein ("BPI"; SEQ ID NO: 38)

protein is shown in Figures 5F and 5G, wherein identical amino acid residues are indicated by ":" and similar amino acid residues are indicated by ".". An alignment of the amino acid sequences of human TANGO 286 ("286"; SEQ ID NO: 35) and RENP protein ("RENP"; SEQ ID NO: 39) is shown in Figures 5H and 5I, wherein identical amino acid residues are indicated by ":" and similar amino acid residues are indicated by ".".

5

10

15

20

25

30

Figure 6 comprises Figures 6A-6H. The nucleotide sequence (SEQ ID NO: 45) of a cDNA encoding the human TANGO 294 protein described herein is listed in Figures 6A-6C. The ORF (residues 126 to 1394; SEQ ID NO: 46) of the cDNA is indicated by nucleotide triplets, above which the amino acid sequence (SEQ ID NO: 47) of human TANGO 294 is listed. An alignment of the amino acid sequences of human TANGO 294 protein ("294"; SEQ ID NO: 47) and a known human lipase protein ("HLP"; SEQ ID NO: 75; GenBank Accession No.

NP_004181) is shown in Figures 6D and 6E, wherein identical amino acid residues are indicated by ":" and similar amino acid residues are indicated by ".". Figure 6F is a hydrophilicity plot of TANGO 294 protein. An alignment of the amino acid sequences of human TANGO 294 protein ("294"; SEQ ID NO: 47) and a known human lysosomal acid lipase protein ("LAL"; SEQ ID NO: 41) is shown in Figures 6G and 6H, wherein identical amino acid residues are indicated by ":" and similar amino acid residues are indicated by ":" and similar amino acid residues are indicated by ":" and similar amino acid residues are indicated by ":" and similar amino acid residues are indicated by ":"

Figure 7 comprises Figures 7A-7F. The nucleotide sequence (SEQ ID NO: 53) of a cDNA encoding the human INTERCEPT 296 protein described herein is listed in Figures 7A-7C. The ORF (residues 70 to 1098; SEQ ID NO: 54) of the cDNA is indicated by nucleotide triplets, above which the amino acid sequence (SEQ ID NO: 55) of human INTERCEPT 296 protein is listed. Figure 7D is a hydrophilicity plot of INTERCEPT 296 protein. An alignment of the amino acid sequences of human INTERCEPT 296 protein ("296"; SEQ ID NO: 55) and C. elegans C06E1.3 related protein ("CRP"; SEQ ID NO: 40) is shown in Figure 7E and 7F, wherein identical amino acid residues are indicated by ":" and similar amino acid residues are indicated by "."

Detailed Description of the Invention

The present invention is based, at least in part, on the discovery of a variety of human and murine cDNA molecules which encode proteins which are herein designated TANGO 202, TANGO 234, TANGO 265, TANGO 273, TANGO 286, TANGO 294, and INTERCEPT 296. These proteins exhibit a variety of physiological activities, and are included in a single application for the sake of convenience. It is understood that the allowability or non-allowability of claims directed to one of these proteins has no bearing on the allowability of claims directed to the others. The characteristics of each of these proteins and the cDNAs encoding them are now described separately.

TANGO 202

5

10

15

20

25

30

A cDNA clone (designated jthke096b05) encoding at least a portion of human TANGO 202 protein was isolated from a human fetal skin cDNA library. The corresponding murine cDNA was isolated as a clone (designated jtmMa044f07) from a bone marrow stromal cell cDNA library. The human TANGO 202 protein is predicted by structural analysis to be a type I membrane protein, although it can exist in a secreted form as well. The murine TANGO 202 protein is predicted by structural analysis to be a secreted protein.

The full length of the cDNA encoding human TANGO 202 protein (Figure 1; SEQ ID NO: 1) is 1656 nucleotide residues. The open reading frame (ORF) of this cDNA, nucleotide residues 34 to 1458 of SEQ ID NO: 1 (i.e., SEQ ID NO: 2), encodes a 475-amino acid transmembrane protein (Figure 1; SEQ ID NO: 3).

The invention thus includes purified human TANGO 202 protein, both in the form of the immature 475 amino acid residue protein (SEQ ID NO: 3) and in the form of the mature 456 amino acid residue protein (SEQ ID NO: 5). The invention also includes purified murine TANGO 202 protein, both in the form of the immature 470 amino acid residue protein (SEQ ID NO: 67) and in the form of the mature 451 amino acid residue protein (SEQ ID NO: 43). Mature human or murine TANGO 202 proteins can be synthesized without the signal sequence

polypeptide at the amino terminus thereof, or they can be synthesized by generating immature TANGO 202 protein and cleaving the signal sequence therefrom.

In addition to full length mature and immature human and murine TANGO 202 proteins, the invention includes fragments, derivatives, and variants of these TANGO 202 proteins, as described herein. These proteins, fragments, derivatives, and variants are collectively referred to herein as polypeptides of the invention or proteins of the invention.

5

10

15

20

25

30

The invention also includes nucleic acid molecules which encode a polypeptide of the invention. Such nucleic acids include, for example, a DNA molecule having the nucleotide sequence listed in SEQ ID NO: 1 or some portion thereof or SEQ ID NO: 67 or some portion thereof, such as the portion which encodes mature human or murine TANGO 202 protein, immature human or murine TANGO 202 protein, or a domain of human or murine TANGO 202 protein. These nucleic acids are collectively referred to as nucleic acids of the invention.

TANGO 202 proteins and nucleic acid molecules encoding them comprise a family of molecules having certain conserved structural and functional features. As used herein, the term "family" is intended to mean two or more proteins or nucleic acid molecules having a common or similar domain structure and having sufficient amino acid or nucleotide sequence identity as defined herein. Family members can be from either the same or different species (e.g., human and mouse, as described herein). For example, a family can comprise two or more proteins of human origin, or can comprise one or more proteins of human origin and one or more of non-human origin.

A common domain present in TANGO 202 proteins is a signal sequence. As used herein, a signal sequence includes a peptide of at least about 10 amino acid residues in length which occurs at the amino terminus of membrane-bound and secreted proteins and which contains at least about 45% hydrophobic amino acid residues such as alanine, leucine, isoleucine, phenylalanine, proline, tyrosine, tryptophan, or valine. In a preferred embodiment, a signal sequence contains at least about 10 to 35 amino acid residues, preferably about 10 to 20 amino acid residues, and has at least about 35-60%, more preferably 40-50%, and more preferably at least about 45% hydrophobic residues. A signal sequence serves

to direct a protein containing such a sequence to a lipid bilayer. Thus, in one embodiment, a TANGO 202 protein contains a signal sequence corresponding to amino acid residues 1 to 19 of SEQ ID NO: 3 (SEQ ID NO: 4) or to amino acid residues 1 to 19 of SEQ ID NO: 69 (SEQ ID NO: 42). The signal sequence is cleaved during processing of the mature protein.

5

10

15

20

25

30

TANGO 202 proteins can also include an extracellular domain. As used herein, an "extracellular domain" refers to a portion of a protein which is localized to the non-cytoplasmic side of a lipid bilayer of a cell when a nucleic acid encoding the protein is expressed in the cell. The human TANGO 202 protein extracellular domain is located from about amino acid residue 20 to about amino acid residue 392 of SEQ ID NO: 3 in the non-secreted form, and from about amino acid residue 20 to amino acid residue 475 of SEQ ID NO: 3 (i.e., the entire mature human protein). The murine TANGO 202 protein extracellular domain is located from about amino acid residue 20 to amino acid residue 470 of SEQ ID NO: 69 (i.e., the entire mature murine protein).

TANGO 202 proteins of the invention can also include a transmembrane domain. As used herein, a "transmembrane domain" refers to an amino acid sequence having at least about 20 to 25 amino acid residues in length and which contains at least about 65-70% hydrophobic amino acid residues such as alanine, leucine, phenylalanine, protein, tyrosine, tryptophan, or valine. In a preferred embodiment, a transmembrane domain contains at least about 15 to 30 amino acid residues, preferably about 20-25 amino acid residues, and has at least about 60-80%, more preferably 65-75%, and more preferably at least about 70% hydrophobic residues. Thus, in one embodiment, a TANGO 202 protein of the invention contains a transmembrane domain corresponding to about amino acid residues 393 to 415 of SEQ ID NO: 3 (SEQ ID NO: 7).

In addition, TANGO 202 proteins of the invention can include a cytoplasmic domain, particularly including a carboxyl-terminal cytoplasmic domain. As used herein, a "cytoplasmic domain" refers to a portion of a protein which is localized to the cytoplasmic side of a lipid bilayer of a cell when a nucleic acid encoding the protein is expressed in the cell. The cytoplasmic domain is

located from about amino acid residue 416 to amino acid residue 475 of SEQ ID NO: 3 (SEQ ID NO: 8) in the non-secreted form of human TANGO 202 protein.

TANGO 202 proteins typically comprise a variety of potential post-translational modification sites (often within an extracellular domain), such as those described herein in Tables I (for human TANGO 202) and II (for murine TANGO 202), as predicted by computerized sequence analysis of TANGO 202 proteins using amino acid sequence comparison software (comparing the amino acid sequence of TANGO 202 with the information in the PROSITE database {rel. 12.2; Feb, 1995} and the Hidden Markov Models database {Rel. PFAM 3.3}).

10

Table I

| Type of Potential Modification Site | Amino Acid Residues | Amino Acid |
|---------------------------------------|---------------------|------------|
| or Domain | of SEQ ID NO: 3 | Sequence |
| N-glycosylation site | 47 to 50 | NWTA |
| | 61 to 64 | NETF |
| | 219 to 222 | NYSA |
| | 295 to 298 | NVSL |
| | 335 to 338 | NQTV |
| | 347 to 350 | NLSV |
| Protein kinase C phosphorylation site | 70 to 72 | TLK |
| | 137 to 139 | TSK |
| | 141 to 143 | SNK |
| | 155 to 157 | SQR |
| | 238 to 240 | TGR |
| | 245 to 247 | TIR |
| | 277 to 279 | THR |
| | 307 to 309 | SDR |
| | 355 to 357 | SSK |
| | 387 to 389 | SHR |
| | 418 to 420 | TFK |
| | 421 to 423 | SHR |

Table I (Continued)

| Casein kinase II phosphorylation site | 337 to 340 | TVAE |
|---------------------------------------|------------|--------------|
| | 438 to 441 | TSGE |
| | 464 to 467 | SQQD |
| N-myristoylation site | 53 to 58 | GGKPCL |
| | 120 to 125 | GNLGCY |
| | 136 to 141 | GTSKTS |
| | 162 to 167 | GMESGY |
| | 214 to 219 | GACGGN |
| Kringle domain signature | 85 to 90 | YCRNPD |
| Kringle Domain | 34 to 116 | · See Fig. 1 |
| CUB domain | 216 to 320 | See Fig. 1 |

Table II

| Type of Potential Modification Site | Amino Acid Residues of | Amino Acid |
|-------------------------------------|------------------------|------------|
| or Domain | SEQ ID NO: 69 | Sequence |
| N-glycosylation site | 59 to 62 | NETF |
| | 217 to 220 | NYSA |
| | 255 to 258 | NFTL |
| · | 293 to 296 | NVSL |
| . * | 333 to 336 | NQTL |
| | 345 to 348 | NLSV |
| cAMP- or cGMP-dependent protein | 455 to 458 | RRSS |
| kinase phosphorylation site | | |

PCT/US00/14858

Table II (Continued)

| Protein kinase C phosphorylation site | 68 to 70 | TLK |
|---------------------------------------|------------|--------|
| Protein kinase C phosphorylation site | | |
| | 135 to 137 | TSK |
| | 139 to 141 | SNK |
| | 153 to 155 | SQR |
| | 236 to 238 | TGR |
| | 243 to 245 | TIR |
| | 275 to 277 | THR |
| | 283 to 285 | SGR |
| | 305 to 307 | SDR |
| * | 353 to 355 | SSK |
| | 408 to 410 | SQR |
| | 453 to 455 | SLR |
| | 457 to 459 | SSR |
| Casein kinase II phosphorylation site | 28 to 31 | SGPE |
| | 257 to 260 | TLFD |
| | 321 to 324 | TKEE |
| | 335 to 338 | TLAE |
| | 384 to 387 | TATE |
| N-myristoylation site | 51 TO 56 | GGKPCL |
| | 118 TO 123 | GNLGCY |
| | 134 TO 139 | GTSKTS |
| | 160 TO 165 | GMESGY |
| | 212 TO 217 | GACGGN |
| | 391 TO 396 | GLCTAW |
| | 429 TO 434 | GTVVSL |
| | | |

Table II (Continued)

| Kringle domain signature | 83 to 88 | YCRNPD |
|--------------------------|------------|------------|
| Kringle Domain | 32 to 114 | See Fig. 1 |
| CUB domain | 214 to 318 | See Fig. 1 |

As used herein, the term "post-translational modification site" refers to a protein domain that includes about 3 to 10 amino acid residues, more preferably about 3 to 6 amino acid residues wherein the domain has an amino acid sequence which comprises a consensus sequence which is recognized and modified by a protein-modifying enzyme. Exemplary protein-modifying enzymes include amino acid glycosylases, cAMP- and cGMP-dependent protein kinases, protein kinase C, casein kinase II, myristoylases, and prenyl transferases. In various embodiments, the protein of the invention has at least 1, 2, 4, 6, 10, 15, or 20 or more of the post-translational modification sites described herein in Tables I and II.

Exemplary additional domains present in human and murine TANGO 202 protein include Kringle domains and CUB domains. In one embodiment, the protein of the invention has at least one domain that is at least 55%, preferably at least about 65%, more preferably at least about 75%, yet more preferably at least about 85%, and most preferably at least about 95% identical to one of the domains described herein in Tables I and II. Preferably, the protein of the invention has at least one Kringle domain and one CUB domain.

A Kringle domain has a characteristic profile that has been described in the art (Castellino and Beals (1987) J. Mol. Evol. 26:358-369; Patthy (1985) Cell 41:657-663; Ikeo et al. (1991) FEBS Lett. 287:146-148). Many, but not all, Kringle domains comprise a conserved hexapeptide signature sequence, namely

$$(F \text{ or } Y) - C - R - N - P - (D \text{ or } N \text{ or } R).$$

25 The cysteine residue is involved in a disulfide bond.

5

10

15

20

Kringle domains are triple-looped, disulfide cross-linked domains found in a varying number of copies in, for example, some serine proteases and plasma proteins. Kringle domains have a role in binding mediators (e.g., membranes, other proteins, or phospholipids) and in regulation of proteolytic

activity. Kringle domains have been identified in the following proteins, for example: apolipoprotein A, blood coagulation factor XII (Hageman factor), hepatocyte growth factor (HGF), HGF-like protein (Friezner Degen et al., (1991) Biochemistry 30:9781-9791), HGF activator (Miyazawa et al., (1993) J. Biol. Chem. 268:10024-10028), plasminogen, thrombin, tissue plasminogen activator, urokinase-type plasminogen activator, and four influenza neuraminidases. The presence of a Kringle domain in each of human and murine TANGO 202 protein indicates that TANGO 202 is involved in one or more physiological processes in which these other Kringle domain-containing proteins are involved, has biological activity in common with one or more of these other Kringle domain-containing proteins, or both.

CUB domains are extracellular domains of about 110 amino acid residues which occur in functionally diverse, mostly developmentally regulated proteins (Bork and Beckmann (1993) J. Mol. Biol. 231:539-545; Bork (1991) FEBS Lett. 282:9-12). Many CUB domains contain four conserved cysteine residues, although some, like that of TANGO 202, contain only two of the conserved cysteine residues. The structure of the CUB domain has been predicted to assume a beta-barrel configuration, similar to that of immunoglobulins. Other proteins which have been found to comprise one or more CUB domains include, for example, mammalian complement sub-components Cls and Clr, hamster serine protease Casp, mammalian complement activating component of Ra-reactive factor, vertebrate enteropeptidase, vertebrate bone morphogenic protein 1, sea urchin blastula proteins BP10 and SpAN, Caenorhabditis elegans hypothetical proteins F42A10.8 and R151.5, neuropilin (A5 antigen), sea urchin fibropellins I and III, mammalian hyaluronate-binding protein TSG-6 (PS4), mammalian spermadhesins, and Xenopus embryonic protein UVS.2. The presence of a CUB domain in each of human and murine TANGO 202 protein indicates that TANGO 202 is involved in one or more physiological processes in which these other CUB domain-containing proteins are involved, has biological activity in common with one or more of these other CUB domain-containing proteins, or both.

15

20

25

30

The signal peptide prediction program SIGNALP (Nielsen et al. (1997) Protein Engineering 10:1-6) predicted that human TANGO 202 protein

includes a 19 amino acid signal peptide (amino acid residues 1 to 19 of SEQ ID NO: 3; SEQ ID NO: 4) preceding the mature TANGO 202 protein (amino acid residues 20 to 475 of SEQ ID NO: 3; SEQ ID NO: 5). Human TANGO 202 protein includes an extracellular domain (amino acid residues 20 to 392 of SEQ ID NO: 3; SEQ ID NO: 6); a transmembrane domain (amino acid residues 393 to 415 of SEQ ID NO: 3; SEQ ID NO: 7); and a cytoplasmic domain (amino acid residues 416 to 475 of SEQ ID NO: 3; SEQ ID NO: 8). The murine homolog of TANGO 202 protein is predicted to be a secreted protein. Thus, it is recognized that human TANGO 202 can also exist in the form of a secreted protein, likely being translated from an alternatively spliced TANGO 202 mRNA. In a variant form of the protein, an extracellular portion of TANGO 202 protein (e.g., amino acid residues 20 to 392 of SEQ ID NO: 3) can be cleaved from the mature protein to generate a soluble fragment of TANGO 202.

5

10

15

20

25

30

Figure 1L depicts a hydrophilicity plot of human TANGO 202 protein. Relatively hydrophobic regions are above the dashed horizontal line, and relatively hydrophilic regions are below the dashed horizontal line. The hydrophobic region which corresponds to amino acid residues 1 to 19 of SEQ ID NO: 3 is the signal sequence of human TANGO 202 (SEQ ID NO: 4). The hydrophobic region which corresponds to amino acid residues 393 to 415 of SEQ ID NO: 3 is the transmembrane domain of human TANGO 202 (SEQ ID NO: 7). As described elsewhere herein, relatively hydrophilic regions are generally located at or near the surface of a protein, and are more frequently effective immunogenic epitopes than are relatively hydrophobic regions. For example, the region of human TANGO 202 protein from about amino acid residue 61 to about amino acid residue 95 appears to be located at or near the surface of the protein, while the region from about amino acid residue 395 to about amino acid residue 420 appears not to be located at or near the surface.

The predicted molecular weight of human TANGO 202 protein without modification and prior to cleavage of the signal sequence is about 51.9 kilodaltons. The predicted molecular weight of the mature human TANGO 202 protein without modification and after cleavage of the signal sequence is about 50.1 kilodaltons.

The full length of the cDNA encoding murine TANGO 202 protein (Figure 1; SEQ ID NO: 67) is 4928 nucleotide residues. The ORF of this cDNA, nucleotide residues 81 to 1490 of SEQ ID NO: 67 (i.e., SEQ ID NO: 68), encodes a 470-amino acid secreted protein (Figure 1; SEQ ID NO: 69).

The signal peptide prediction program SIGNALP (Nielsen et al. (1997) Protein Engineering 10:1-6) predicted that murine TANGO 202 protein includes a 19 amino acid signal peptide (amino acid residues 1 to 19 of SEQ ID NO: 69; SEQ ID NO: 42) preceding the mature TANGO 202 protein (amino acid residues 20 to 470 of SEQ ID NO: 69; SEQ ID NO: 43). Murine TANGO 202 protein is a secreted protein.

5

10

15

20

25

30

protein. Relatively hydrophobic regions are above the dashed horizontal line, and relatively hydrophilic regions are below the dashed horizontal line. The hydrophobic region which corresponds to amino acid residues 1 to 19 of SEQ ID NO: 69 is the signal sequence of murine TANGO 202 (SEQ ID NO: 42). As described elsewhere herein, relatively hydrophilic regions are generally located at or near the surface of a protein, and are more frequently effective immunogenic epitopes than are relatively hydrophobic regions. For example, the region of murine TANGO 202 protein from about amino acid residue 61 to about amino acid residue 95 appears to be located at or near the surface of the protein, while the region from about amino acid residue 295 to about amino acid residue 305 appears not to be located at or near the surface

The predicted molecular weight of murine TANGO 202 protein without modification and prior to cleavage of the signal sequence is about 51.5 kilodaltons. The predicted molecular weight of the mature murine TANGO 202 protein without modification and after cleavage of the signal sequence is about 49.7 kilodaltons.

Human and murine TANGO 202 proteins exhibit considerable sequence similarity, as indicated herein in Figures 1J and 1K. Figures 1J and 1K depict an alignment of human and murine TANGO 202 amino acid sequences (SEQ ID NOs: 3 and 69, respectively). In this alignment (made using the ALIGN software {Myers and Miller (1989) CABIOS, ver. 2.0}; pam120.mat scoring matrix;

gap penalties -12/-4), the proteins are 76.5% identical. The human and murine ORFs encoding TANGO 202 are 87.4% identical, as assessed using the same software and parameters.

In situ hybridization experiments in mouse tissues indicated that mRNA corresponding to the cDNA encoding TANGO 202 is expressed in the tissues listed in Table III, wherein "+" indicates detectable expression and "++" indicates a greater level of expression than "+".

Table III

| Animal | Tissue | Relative Level of Expression |
|-------------------|--|---------------------------------|
| Mouse (Adult) | bladder, especially in transitional epithelium | ++ |
| | renal glomeruli | + |
| | brain | + |
| | heart | + |
| | liver | + |
| | spleen | + |
| · | placenta | + |
| Mouse (Embryo) | ubiquitous | + |

10

15

5

Biological function of TANGO 202 proteins, nucleic acids, and modulators thereof

TANGO 202 proteins are involved in disorders which affect both tissues in which they are normally expressed and tissues in which they are normally not expressed. Based on the observation that TANGO 202 is expressed in human fetal skin, ubiquitously in fetal mouse tissues, in adult murine bone marrow stromal cells, and in cells of adult murine bladder, renal glomeruli, brain, heart, liver, spleen and placenta, TANGO 202 protein is involved in one or more biological processes

which occur in these tissues. In particular, TANGO 202 is involved in modulating growth, proliferation, survival, differentiation, and activity of cells of these tissues including, but not limited to, hematopoietic and fetal cells. Thus, TANGO 202 has a role in disorders which affect these cells and their growth, proliferation, survival, differentiation, and activity. Ubiquitous expression of TANGO 202 in fetal murine tissues, contrasted with limited expression in adult murine tissues further indicates that TANGO 202 is involved in disorders in which it is inappropriately expressed (e.g., disorders in which TANGO 202 is expressed in adult murine tissues other than bone marrow stromal cells and disorders in which TANGO 202 is not expressed in one or more developing fetal tissues).

5

10

15

20

25

30

The presence of a Kringle domain in both the murine and human TANGO 202 proteins indicates that this protein is involved in modulating cellular binding to one or more mediators (e.g., proteins, phospholipids, intracellular organelles, or other cells), in modulating proteolytic activity, or both. The presence of a Kringle domain in other proteins (e.g., growth factors) indicates activities that these proteins share with TANGO 202 protein (e.g., modulating cell dissociation and migration into and through extracellular matrices). The presence of Kringle domains in numerous plasma proteins, particularly coupled with the observation that TANGO 202 is expressed in adult murine bone marrow stromal cells, indicates a role for TANGO 202 protein in modulating binding of blood or hematopoietic cells (or both) to one or more mediators. Thus, TANGO 202 is involved in disorders relating to aberrant cellular protease activity, inappropriate interaction or non-interaction of cells with mediators, and in blood and hematopoietic cell-related disorders. Such disorders include, by way of example and not limitation, immune disorders, infectious diseases, auto-immune disorders, vascular and cardiovascular disorders, disorders related to mal-expression of growth factors, cancers, hematological disorders, and the like.

The cDNA encoding TANGO 202 exhibits significant nucleotide sequence similarity with a polynucleotide encoding a kringle-domain-containing protein (designated HTHBZ47) described in the European Patent Application No. EP 0 911 399 A2 (published April 28, 1999). Thus, the TANGO 202 protein can exhibit one or more of the activities exhibited by HTHBZ47, and can be used to

prevent, inhibit, diagnose, and treat one or more disorders for which HTHBZ47 is useful. These disorders include cancer, inflammation, autoimmune disorders, allergic disorders, asthma, rheumatoid arthritis, inflammation of central nervous system tissues, cerebellar degeneration, Alzheimer's disease, Parkinson's disease, multiple sclerosis, amylotrophic lateral sclerosis, head injury damage and other neurological abnormalities, septic shock, sepsis, stroke, osteoporosis, osteoarthritis, ischemic reperfusion injury, cardiovascular disease, kidney disease, liver disease, ischemic injury, myocardial infarction, hypotension, hypertension, AIDS, myelodysplastic syndromes and other hematologic abnormalities, aplastic anemia, male pattern baldness, and bacterial, fungal, protozoan, and viral infections.

5

10

15

20

25

30

The presence of a CUB domain in both the murine and human TANGO 202 proteins indicates that this protein is involved in biological processes common to other CUB domain-containing proteins, such as developmental processes and binding to mediators. Therefore, TANGO 202 protein has a role in disorders which involve inappropriate developmental processes (e.g., abnormally high proliferation or un-differentiation of a differentiated tissue or abnormally low differentiation or proliferation of a non-developed or non-differentiated tissue) and modulation of cell growth, proliferation, survival, differentiation, and activity. Such disorders include, by way of example and not limitation, various cancers and birth and developmental defects.

Thus, proteins and nucleic acids of the invention which are identical to, similar to, or derived from human and murine TANGO 202 proteins and nucleic acids encoding them are useful for preventing, diagnosing, and treating, among others, vascular and cardiovascular disorders, hematological disorders, disorders related to mal-expression of growth factors, and cancer. Other uses for these proteins and nucleic acids of the invention relate to modulating cell growth (e.g., angiogenesis), proliferation (e.g., cancers), survival (e.g., apoptosis), differentiation (e.g., hematopoiesis), and activity (e.g., ligand-binding capacity). TANGO 202 proteins and nucleic acids encoding them are also useful for modulating cell dissociation and modulating migration of cells in extracellular matrices.

TANGO 234

5

10

15

20

25

30

A cDNA clone (designated jthsa104d11) encoding at least a portion of human TANGO 234 protein was isolated from a human fetal spleen cDNA library. The human TANGO 234 protein is predicted by structural analysis to be a transmembrane protein, although it can exist in a secreted form as well.

The full length of the cDNA encoding human TANGO 234 protein (Figure 2; SEQ ID NO: 9) is 4628 nucleotide residues. The ORF of this cDNA, nucleotide residues 28 to 4386 of SEQ ID NO: 9 (i.e., SEQ ID NO: 10), encodes a 1453-amino acid transmembrane protein (Figure 2; SEQ ID NO: 11).

The invention thus includes purified human TANGO 234 protein, both in the form of the immature 1453 amino acid residue protein (SEQ ID NO: 11) and in the form of the mature 1413 amino acid residue protein (SEQ ID NO: 13). Mature human TANGO 234 protein can be synthesized without the signal sequence polypeptide at the amino terminus thereof, or it can be synthesized by generating immature TANGO 234 protein and cleaving the signal sequence therefrom.

In addition to full length mature and immature human TANGO 234 proteins, the invention includes fragments, derivatives, and variants of these TANGO 234 proteins, as described herein. These proteins, fragments, derivatives, and variants are collectively referred to herein as polypeptides of the invention or proteins of the invention.

The invention also includes nucleic acid molecules which encode a polypeptide of the invention. Such nucleic acids include, for example, a DNA molecule having the nucleotide sequence listed in SEQ ID NO: 9 or some portion thereof, such as the portion which encodes mature TANGO 234 protein, immature TANGO 234 protein, or a domain of TANGO 234 protein. These nucleic acids are collectively referred to as nucleic acids of the invention.

TANGO 234 proteins and nucleic acid molecules encoding them comprise a family of molecules having certain conserved structural and functional features, as indicated by the conservation of amino acid sequence between human TANGO 234 protein and bovine WC1 protein, as shown in Figures 2K through 2P, and the conservation of nucleotide sequence between the ORFs encoding human

TANGO 234 protein and bovine WC1 protein, as shown in Figures 2Qi through 2Qxvii.

5

10

15

20

25

30

A common domain present in TANGO 234 proteins is a signal sequence. As used herein, a signal sequence includes a peptide of at least about 10 amino acid residues in length which occurs at the amino terminus of membrane-bound proteins and which contains at least about 45% hydrophobic amino acid residues such as alanine, leucine, isoleucine, phenylalanine, proline, tyrosine, tryptophan, or valine. In a preferred embodiment, a signal sequence contains at least about 10 to 35 amino acid residues, preferably about 10 to 20 amino acid residues, and has at least about 35-60%, more preferably 40-50%, and more preferably at least about 45% hydrophobic residues. A signal sequence serves to direct a protein containing such a sequence to a lipid bilayer. Thus, in one embodiment, a TANGO 234 protein contains a signal sequence corresponding to amino acid residues 1 to 40 of SEQ ID NO: 11 (SEQ ID NO: 12). The signal sequence is cleaved during processing of the mature protein.

TANGO 234 proteins can include an extracellular domain. The human TANGO 234 protein extracellular domain is located from about amino acid residue 41 to about amino acid residue 1359 of SEQ ID NO: 3. TANGO 234 can alternately exist in a secreted form, such as a mature protein having the amino acid sequence of amino acid residues 41 to 1453 or residues 41 to about 1359 of SEQ ID NO: 11.

In addition, TANGO 234 include a transmembrane domain. In one embodiment, a TANGO 234 protein of the invention contains a transmembrane domain corresponding to about amino acid residues 1360 to 1383 of SEQ ID NO: 11 (SEQ ID NO: 15).

The present invention includes TANGO 234 proteins having a cytoplasmic domain, particularly including proteins having a carboxyl-terminal cytoplasmic domain. The human TANGO 234 cytoplasmic domain is located from about amino acid residue 1384 to amino acid residue 1453 of SEQ ID NO: 11 (SEQ ID NO: 16).

TANGO 234 proteins typically comprise a variety of potential posttranslational modification sites (often within an extracellular domain), such as those

described herein in Table IV, as predicted by computerized sequence analysis of TANGO 234 proteins using amino acid sequence comparison software (comparing the amino acid sequence of TANGO 234 with the information in the PROSITE database {rel. 12.2; Feb, 1995} and the Hidden Markov Models database {Rel. PFAM 3.3}). In certain embodiments, a protein of the invention has at least 1, 2, 4, 6, 10, 15, or 20 or more of the post-translational modification sites listed in Table IV.

Table IV

| Type of Potential Modification Site | Amino Acid Residues of | Amino Acid |
|-------------------------------------|------------------------|------------|
| or Domain | SEQ ID NO: 11 | Sequence |
| N-glycosylation site | 42 to 45 | NGTD |
| | 78 to 81 | NTTA |
| | 120 to 123 | NESA |
| | 161 to 164 | NNSC |
| , , | 334 to 337 | NESF |
| | 377 to 380 | NCSG |
| | 441 to 444 | NESA |
| | 548 to 551 | NESN |
| | 637 to 640 | NAST |
| • | 972 to 975 | NESL |
| | 1013 to 1016 | NVSD |
| * | 1084 to 1087 | NATV |
| | 1104 to 1107 | NCTG |
| · | 1161 to 1164 | NGTW |
| | 1171 to 1174 | NITT |
| . · · · · · | 1318 to 1321 | NESF |
| | 1354 to 1357 | NASS |
| Glycosaminoglycan attachment site | 558 to 561 | SGWG |
| | 665 to 668 | SGWG |
| cAMP- or cGMP-dependent protein | 1229 to 1232 | RRIS |
| kinase phosphorylation site | 1399 to 1402 | RRGS |

Table IV (Continued)

| Protein kinase C phosphorylation site | 165 to 167 | SGR |
|---------------------------------------|--------------|------------------|
| Protein kinase C phosphoryladion site | 268 to 270 | TNR |
| | | SGR |
| | 379 to 381 | |
| | 419 to 421 | SRR |
| | 469 to 471 | SDK |
| | 506 to 508 | STR |
| | 589 to 591 | SNR [.] |
| | 593 to 595 | SGR |
| | 661 to 663 | SCR |
| | 696 to 698 | SSR |
| _ | 746 to 748 | TER |
| · | 805 to 807 | SGR |
| | 815 to 817 | TWR |
| | 959 to 961 | SVR |
| | 1256 to 1258 | SGR |
| | 1349 to 1351 | SLK · |
| | 1396 to 1398 | STR |
| Casein kinase II phosphorylation site | 44 to 47 | TDLE |
| | 71 to 74 | TVCD |
| · | 178 to 181 | TICD |
| | 245 to 248 | SHNE |
| | 253 to 256 | TCYD |
| | 258 to 261 | SDLE |
| | 319 to 322 | SGSD |
| | 332 to 335 | SGNE |
| | 392 to 395 | TICD |
| a * . | 439 to 442 | TGNE |

Table IV (Continued)

| Casein kinase II phosphorylation site | 606 to 609 | TVCD |
|---------------------------------------|--------------|--------|
| (Continued) | 622 to 625 | SQLD |
| | 673 to 676 | SHSE |
| | 686 to 689 | SDME |
| | 760 to 763 | TGGE |
| | 765 to 768 | SLWD |
| | 818 to 821 | SVCD |
| | 845 to 848 | SVGD |
| ~ | 857 to 860 | TWAE |
| | 907 to 910 | SQCD |
| | 923 to 926 | SLCD |
| <u> </u> | 927 to 930 | THWD |
| | 974 to 977 | SLLD |
| | 1059 to 1062 | TICD |
| | 1106 to 1109 | TGTE |
| | 1145 to 1148 | SETE |
| | 1233 to 1236 | SPAE |
| · | 1241 to 1244 | TCED |
| | 1269 to 1272 | TVCD |
| | 1402 to 1405 | SLEE |
| | 1425 to 1428 | TSDD |
| N-myristoylation site | 67 to 72 | GQWGTV |
| | 90 to 95 | GCPFSF |
| | 101 to 106 | GQAVTR |
| | 119 to 124 | GNESAL |
| | 133 to 138 | GSHNCY |
| · | 160 to 165 | GNNSCS |
| - | 197 to 202 | GCPSSF |

Table IV (Continued)

| N-myristoylation site (Continued) | 226 to 231 | GNELAL |
|-----------------------------------|--------------|----------|
| | 240 to 245 | GNHDCS |
| · | 267 to 272 | GTNRCM |
| | 304 to 309 | GCGTAL |
| | 328 to 333 | GVSCSG |
| | 374 to 379 | GSNNCS |
| | 411 to 416 | GCPFSV |
| | 418 to 423 | GSRRAK |
| | 440 to 445 | GNESAL |
| | 465 to 470 | GVICSD |
| | 547 to 552 | GNESNI |
| • | 588 to 593 | GSNRCS |
| | 632 to 637 | GMGLGN |
| | 668 to 673 | GNNDCS |
| | 679 to 684 | GVICSD |
| | 695 to 700 | - GSSRCA |
| | 712 to 717 | GILCAN |
| | 720 to 725 | GMNIAE |
| | 758 to 763 | GCTGGE |
| | 853 to 858 | GNGLTW |
| | 891 to 896 | GVVCSR |
| | 944 to 949 | GTALST |
| | 985 to 990 | GAPPCI |
| | 992 to 997 | GNTVSV |
| | 1078 to 1083 | GCGVAF |
| | 1121 to 1126 | GQHDCR |
| · | 1132 to 1137 | GVICSE |

Table IV (Continued)

| N 11 11 11 11 11 11 11 11 11 11 11 11 11 | 1162 to 1167 | GTWGSV |
|--|--------------|------------|
| N-myristoylation site (Continued) | | |
| , | 1185 to 1190 | GCGENG |
| | 1265 to 1270 | GSWGTV |
| | 1288 to 1293 | GCGSAL |
| | 1302 to 1307 | GQGTGT |
| | 1331 to 1336 | GQSDCG |
| | 1342 to 1347 | GVRCSG |
| | 1422 to 1427 | GTRTSD |
| | 1443 to 1438 | GCEDAS |
| | 1444 to 1449 | GVLPAS |
| Amidation site | 1167 to 1170 | VGRR |
| Speract receptor repeated (SRR) | 53 to 90 | See Fig. 2 |
| domain signature | 160 to 197 | See Fig. 2 |
| | 267 to 304 | See Fig. 2 |
| | 1041 to 1078 | See Fig. 2 |
| | 1251 to 1288 | See Fig. 2 |
| Scavenger receptor cysteine-rich | 51 to 148 | See Fig. 2 |
| (SRCR) domain | 158 to 255 | See Fig. 2 |
| | 265 to 362 | See Fig. 2 |
| | 372 to 469 | See Fig. 2 |
| | 479 to 576 | See Fig. 2 |
| * | 586 to 683 | See Fig. 2 |
| | 693 to 790 | See Fig. 2 |
| | 798 to 895 | See Fig. 2 |
| * | 903 to 1000 | See Fig. 2 |
| | 1039 to 1136 | See Fig. 2 |
| <i>\$</i> | 1146 to 1243 | See Fig. 2 |
| | 1249 to 1346 | See Fig. 2 |
| | | |

Among the domains that occur in TANGO 234 protein are SRR domains and SRCR domains. In one embodiment, the protein of the invention has at least one domain that is at least 55%, preferably at least about 65%, more preferably at least about 75%, yet more preferably at least about 85%, and most preferably at least about 95% identical to one of these domains. In other embodiments, the protein has at least two of the SRR and SRCR domains described herein in Table IV. In other embodiments, the protein has at least one SRR domain and at least one SRCR domain.

5

10

15

20

25

30

The SRR domain is named after a receptor domain identified in a sea urchin egg protein designated speract. The consensus sequence of this domain (using standard one-letter amino acid codes, wherein X is any amino acid residue) is as follows.

-G-X₃-G-X₂-E-X₆-W-G-X₂-C-X₃-(F or Y or W)-X₈-C-X₃-G-.

Speract is a transmembrane glycoprotein of 500 amino acid residues (Dangott et al. (1989) Proc. Natl. Acad. Sci. USA 86:2128-2132). Structurally, this receptor consists of a large extracellular domain of 450 residues, followed by a transmembrane region and a small cytoplasmic domain of 12 amino acid residues. The extracellular domain contains four repeats of an approximately 115 amino acid domain. There are 17 amino acid residues that are perfectly conserved in the four repeats in speract, including six cysteine residues, six glycine residues, and two glutamate residues. TANGO 234 has five SRR domains, in which 16 of the 17 conserved speract residues are present of four of the SRR domains and 15 are present in the remaining SRR domain. This domain is designated the speract receptor repeated domain. The amino acid sequence of mammalian macrophage scavenger receptor type I (MSRI) exhibits such a domain (Freeman et al. (1990) Proc. Natl. Acad. Sci. USA 87:8810-8814). MSRI proteins are membrane glycoproteins implicated in the pathologic deposition of cholesterol in arterial walls during atherogenesis. TANGO 234 is involved in one or more physiological processes related to cholesterol deposition and atherogenesis, as well as other vascular and cardiovascular disorders.

Scavenger receptor cysteine-rich (SRCR) domains are disulfide rich extracellular domains which are present in certain cell surface and secreted proteins.

Proteins having SRCR domains exhibit diverse ligand binding specificity. For example, in addition to modified lipoproteins, some of these proteins bind a variety of surface components of pathogenic microorganisms, and some of the proteins bind apoptotic cells. SRCR domains are also involved in mediating immune development and response. Other SRCR-containing proteins are involved in binding of modified lipoproteins (e.g., oxidized low density lipoprotein {LDL}) by specialized macrophages, leading to the formation of macrophages filled with cholesteryl ester droplets (i.e., foam cells). TANGO 234 is involved in one or more physiological processes in which these other SRCR domain-containing proteins are involved, such as LDL uptake and metabolism, regulation of serum cholesterol level, atherogenesis, atherosclerosis, bacterial or viral infections, immune development, and generation and perseverance of immune responses.

5

10

15

20

25

30

WC1 is a ruminant protein having an SRCR domain. WC1 and gamma delta T-cell receptor are the only known gamma delta T-cell specific antigens. Antibodies which bind specifically with WC1 induce growth arrest in IL-2-dependent gamma delta T-cell and augment proliferation of gamma delta T-cells in an autologous mixed lymphocyte reaction or in the presence of anti-CD2 or anti-CD5 antibodies. Injection of antibodies which bind specifically with WC1 into calves results in long-lasting depletion of gamma delta T-cells. Furthermore, antibodies which bind specifically with WC1 can be used to purify gamma delta T-cells.

Gamma delta T-cells are involved in a variety of physiological processes. For example, these cells are potential mediators of allergic airway inflammation and lyme disease. Furthermore, these cells are involved in natural resistance to viral infections and can mediate autoimmune diseases. Elimination of gamma delta T-cells by injection of antibodies which bind specifically therewith can affect the outcomes of these disorders.

TANGO 234 is likely the human orthologue of ruminant protein WC1, and thus is involved with the physiological processes described above in humans. An alignment of the amino acid sequences of (human) TANGO 234 and bovine WC1 protein is shown in Figures 2K-2P. In this alignment (made using the ALIGN software {Myers and Miller (1989) CABIOS, ver. 2.0}; pam120.mat

scoring matrix; gap penalties -12/-4), the proteins are 40.4% identical. An alignment of the nucleotide sequences of the ORFs encoding (human) TANGO 234 and bovine WC1 protein is shown in Figures 2Qi-2Qxvii. The two ORFs are 54.3% identical, as assessed using the same software and parameters.

5

10

15

20

25

30

The signal peptide prediction program SIGNALP (Nielsen et al. (1997) Protein Engineering 10:1-6) predicted that human TANGO 234 protein includes a 40 amino acid signal peptide (amino acid residues 1 to 40 of SEQ ID NO: 11; SEQ ID NO: 12) preceding the mature TANGO 234 protein (amino acid residues 41 to 4386 of SEQ ID NO: 11; SEQ ID NO: 13). Human TANGO 234 protein includes an extracellular domain (amino acid residues 41 to 1359 of SEQ ID NO: 11; SEQ ID NO: 14); a transmembrane domain (amino acid residues 1360 to 1383 of SEQ ID NO: 11; SEQ ID NO: 15); and a cytoplasmic domain (amino acid residues 1384 to 1453 of SEQ ID NO: 11; SEQ ID NO: 16).

Figure 2J depicts a hydrophilicity plot of human TANGO 234 protein. Relatively hydrophobic regions are above the dashed horizontal line, and relatively hydrophilic regions are below the dashed horizontal line. The hydrophobic region which corresponds to amino acid residues 1 to 40 of SEQ ID NO: 11 is the signal sequence of human TANGO 234 (SEQ ID NO: 12). The hydrophobic region which corresponds to amino acid residues 1360 to 1383 of SEQ ID NO: 11 is the transmembrane domain of human TANGO 234 (SEQ ID NO: 15). As described elsewhere herein, relatively hydrophilic regions are generally located at or near the surface of a protein, and are more frequently effective immunogenic epitopes than are relatively hydrophobic regions. For example, the region of human TANGO 234 protein from about amino acid residue 225 to about amino acid residue 250 appears to be located at or near the surface of the protein, while the region from about amino acid residue 990 to about amino acid residue 1000 appears not to be located at or near the surface.

The predicted molecular weight of human TANGO 234 protein without modification and prior to cleavage of the signal sequence is about 159.3 kilodaltons. The predicted molecular weight of the mature human TANGO 234 protein without modification and after cleavage of the signal sequence is about 154.7 kilodaltons.

Chromosomal mapping to identify the location of the gene encoding human TANGO 234 protein indicated that the gene was located at chromosomal location h12p13 (with synteny to mo6). Flanking chromosomal markers include WI-6980 and GATA8A09.43. Nearby human loci include IBD2 (inflammatory bowel disease 2), FPF (familial periodic fever), and HPDR2 (hypophosphatemia vitamin D resistant rickets 2). Nearby genes are KLRC (killer cell receptor cluster), DRPLA (dentatorubro-pallidoluysian atrophy), GAPD (glyceraldehyde-3-phosphate) dehydrogenase, and PXR1 (peroxisome receptor 1). Murine chromosomal mapping indicated that the murine orthologue is located near the scr (scruffy) locus. Nearby mouse genes include drpla (dentatorubral phillidoluysian atrophy), prp (proline rich protein), and kap (kidney androgen regulated protein).

5

10

15

20

25

Northern analysis experiments indicated that mRNA corresponding to the cDNA encoding TANGO 234 is expressed in the tissues listed in Table V, wherein "++" indicates moderate expression, "+" indicates lower expression, and "-" indicates no detectable expression.

Table V

| Animal | Tissue | Relative Level of Expression |
|--------|-----------------------------|------------------------------|
| Human | spleen | ++ |
| | fetal lung | ++ |
| | lung | + |
| | thymus | + ' |
| | bone marrow | - |
| ≘. | peripheral blood leukocytes | - |

Biological function of TANGO 234 proteins, nucleic acids, and modulators thereof

TANGO 234 proteins are involved in disorders which affect both tissues in which they are normally expressed and tissues in which they are normally not expressed. Based on the observation that TANGO 234 is expressed in human fetal lung, spleen, and, to a lesser extent in adult lung and thymus tissue, TANGO 234 protein is involved in one or more biological processes which occur in these tissues. In particular, TANGO 234 is involved in modulating growth, proliferation, survival, differentiation, and activity of cells including, but not limited to, lung,

spleen, thymus bone marrow, hematopoietic, peripheral blood leukocytes, and fetal cells of the animal in which it is normally expressed. Thus, TANGO 234 has a role in disorders which affect these cells and their growth, proliferation, survival, differentiation, and activity. Expression of TANGO 234 in an animal is also involved in modulating growth, proliferation, survival, differentiation, and activity of cells and viruses which are foreign to the host (i.e., bacterial, fungal, and viral infections).

Homology of human TANGO 234 with bovine WC1 protein indicates that TANGO 234 has physiological functions in humans analogous to the functions of WC1 in ruminants. Thus, TANGO 234 is involved in modulating growth, proliferation, survival, differentiation, and activity of gamma delta T cells. For example, TANGO 234 affects the ability of gamma delta T cells to interact with chemokines such as interleukin-2. TANGO 234 therefore is involved in the physiological processes associated with allergic airway inflammation, lyme arthritis, resistance to viral infection, auto-immune diseases, and the like.

In addition, presence in TANGO 234 of SRR and SRCR domains indicates that TANGO 234 is involved in physiological functions identical or analogous to the functions performed by other proteins having such domains. For example, like other SRR domain-containing proteins, TANGO 234 modulates cholesterol deposition in arterial walls, and is thus involved in development and persistence of atherogenesis and arteriosclerosis, as well as other vascular and cardiovascular disorders. Like other SRCR domain-containing proteins, TANGO 234 is involved in uptake and metabolism of LDL, regulation of serum cholesterol level, and can modulate these processes as well as the processes of atherogenesis, arteriosclerosis, immune development, and generation and perseverance of immune responses to bacterial, fungal, and viral infections.

TANGO 265

5

10

15

20

25

A cDNA clone (designated jthsa079g01) encoding at least a portion
of human TANGO 265 protein was isolated from a human fetal spleen cDNA
library. The human TANGO 265 protein is predicted by structural analysis to be a
transmembrane membrane protein, although it can exist in a secreted form as well.

The full length of the cDNA encoding human TANGO 265 protein (Figure 3; SEQ ID NO: 17) is 3104 nucleotide residues. The ORF of this cDNA, nucleotide residues 32 to 2314 of SEQ ID NO: 17 (i.e., SEQ ID NO: 18), encodes a 761-amino acid transmembrane protein (Figure 3; SEQ ID NO: 19).

The invention thus includes purified TANGO 265 protein, both in the form of the immature 761 amino acid residue protein (SEQ ID NO: 19) and in the form of the mature 730 amino acid residue protein (SEQ ID NO: 21). Mature TANGO 265 protein can be synthesized without the signal sequence polypeptide at the amino terminus thereof, or it can be synthesized by generating immature TANGO 265 protein and cleaving the signal sequence therefrom.

5

10

15

20

25

30

In addition to full length mature and immature TANGO 265 proteins, the invention includes fragments, derivatives, and variants of TANGO 265 protein, as described herein. These proteins, fragments, derivatives, and variants are collectively referred to herein as polypeptides of the invention or proteins of the invention.

The invention also includes nucleic acid molecules which encode a polypeptide of the invention. Such nucleic acids include, for example, a DNA molecule having the nucleotide sequence listed in SEQ ID NO: 17 or some portion thereof, such as the portion which encodes mature TANGO 265 protein, immature TANGO 265 protein, or a domain of TANGO 265 protein. These nucleic acids are collectively referred to as nucleic acids of the invention.

TANGO 265 proteins and nucleic acid molecules encoding them comprise a family of molecules having certain conserved structural and functional features.

A common domain present in TANGO 265 proteins is a signal sequence. As used herein, a signal sequence includes a peptide of at least about 10 amino acid residues in length which occurs at the amino terminus of membrane-bound proteins and which contains at least about 45% hydrophobic amino acid residues such as alanine, leucine, isoleucine, phenylalanine, proline, tyrosine, tryptophan, or valine. In a preferred embodiment, a signal sequence contains at least about 10 to 35 amino acid residues, preferably about 10 to 20 amino acid residues, and has at least about 35-60%, more preferably 40-50%, and more

preferably at least about 45% hydrophobic residues. A signal sequence serves to direct a protein containing such a sequence to a lipid bilayer. Thus, in one embodiment, a TANGO 265 protein contains a signal sequence corresponding to amino acid residues 1 to 31 of SEQ ID NO: 19 (SEQ ID NO: 20). The signal sequence is cleaved during processing of the mature protein.

5

10

15

20

25

TANGO 265 proteins can also include an extracellular domain. The human TANGO 265 protein extracellular domain is located from about amino acid residue 32 to about amino acid residue 683 of SEQ ID NO: 17. TANGO 265 can alternately exist in a secreted form, such as a mature protein having the amino acid sequence of amino acid residues 32 to 761 or residues 32 to about 683 of SEQ ID NO: 19.

TANGO 265 proteins can also include a transmembrane domain. In one embodiment, a TANGO 265 protein of the invention contains a transmembrane domain corresponding to about amino acid residues 684 to 704 of SEQ ID NO: 19 (SEQ ID NO: 23).

In addition, TANGO 265 proteins include a cytoplasmic domain, particularly including proteins having a carboxyl-terminal cytoplasmic domain. The human TANGO 265 cytoplasmic domain is located from about amino acid residue 705 to amino acid residue 761 of SEQ ID NO: 19 (SEQ ID NO: 24).

TANGO 265 proteins typically comprise a variety of potential post-translational modification sites (often within an extracellular domain), such as those described herein in Table VI, as predicted by computerized sequence analysis of TANGO 265 proteins using amino acid sequence comparison software (comparing the amino acid sequence of TANGO 265 with the information in the PROSITE database {rel. 12.2; Feb, 1995} and the Hidden Markov Models database {Rel. PFAM 3.3}). In certain embodiments, a protein of the invention has at least 1, 2, 4, 6, 10, 15, or 20 or more of the post-translational modification sites listed in Table VI.

Table VI

| Type of Potential Modification Site | Amino Acid Residues | Amino Acid |
|---------------------------------------|---------------------|------------|
| or Domain | of SEQ ID NO: 19 | Sequence |
| | | |
| N-glycosylation site | 120 to 123 | NETQ |
| | 135 to 138 | NVTH |
| | 496 to 499 | NCSV |
| , | 607 to 610 | NGLS |
| Glycosaminoglycan attachment site | 70 to 73 | SGDG |
| cAMP- or cGMP-dependent protein | 108 to 111 | RKKS |
| kinase phosphorylation site | 116 to 119 | KKKS |
| | 281 to 284 | KKWT |
| Protein kinase C phosphorylation site | 106 to 108 | SDR |
| | 262 to 264 | TSR |
| | 361 to 363 | TSR |
| | 366 to 368 | TYR |
| | 385 to 387 | SDK |
| | 533 to 535 | SWK |
| | 555 to 557 | SLR |
| | 721 to 723 | TLR |
| | 738 to 740 | SPK |
| Casein kinase II phosphorylation site | 152 to 155 | TFIE |
| | 176 to 179 | SPFD |
| | 250 to 253 | TASE |
| | 342 to 345 | SLLD |
| | 411 to 414 | SGVE |
| | 498 to 501 | SVYE |
| | 502 to 505 | SCVD |
| | 574 to 577 | SILE |
| , | 738 to 741 | SPKE |
| | 745 to 748 | SASD |

Table VI (Continued)

| N-myristoylation site | 79 to 84 | GAREAI |
|-----------------------|------------|------------|
| • | 191 to 196 | GMLYSG |
| | 331 to 336 | GGTRSS |
| | 412 to 417 | GVEYTR |
| | 437 to 442 | GTTTGS |
| | 620 to 625 | GLYQCW |
| | 671 to 676 | GAALAA |
| Sema domain | 64 to 478 | See Fig. 3 |

An exemplary domains which occurs in TANGO 265 proteins is a sema domain. In one embodiment, the protein of the invention has at least one domain that is at least 55%, preferably at least about 65%, more preferably at least about 75%, yet more preferably at least about 85%, and most preferably at least about 95% identical to one of the sema domains described herein in Table VI.

5

10

15

Sema domains occur in semaphorin proteins. Semaphorins are a large family of secreted and transmembrane proteins, some of which function as repellent signals during neural axon guidance. The sema domain and a variety of semaphorin proteins in which it occurs are described, for example, in Winberg et al. (1998 Cell 95:903-916). Sema domains also occur in human hepatocyte growth factor receptor (Swissprot Accession no. P08581) and the similar neuronal and epithelial transmembrane receptor protein (Swissprot Accession no. P51805). The presence of an sema domain in human TANGO 265 protein indicates that TANGO 265 is involved in one or more physiological processes in which the semaphorins are involved, has biological activity in common with one or more of the semaphorins, or both.

Human TANGO 265 protein exhibits considerable sequence

20 similarity to murine semaphorin B protein (GenBank Accession no. X85991), as
indicated herein in Figures 3F-3H. Figures 3F-3H depict an alignment of the amino
acid sequences of human TANGO 265 protein (SEQ ID NO: 19) and murine
semaphorin B protein (SEQ ID NO: 76). In this alignment (pam120.mat scoring
matrix, gap penalties -12/-4), the amino acid sequences of the proteins are 82.3%

identical. Figures 3I through 3T depict an alignment of the nucleotide sequences of cDNA encoding human TANGO 265 protein (SEQ ID NO: 17) and murine cDNA encoding semaphorin B protein (SEQ ID NO: 77). In this alignment (pam120.mat scoring matrix, gap penalties -12/-4), the nucleic acid sequences of the cDNAs are 76.2% identical. Thus, TANGO 265 is the human orthologue of murine semaphorin B and shares functional similarities to that protein.

It is known that semaphorins are bi-functional, capable of functioning either as attractive axonal guidance proteins or as repellent axonal guidance proteins (Wong et al. (1997) *Development* 124:3597-3607). Furthermore, semaphorins bind with neuronal cell surface proteins designated plexins, which are expressed on both neuronal cells and cells of the immune system (Comeau et al. (1998) *Immunity* 8:473-482; Jin and Strittmatter (1997) *J. Neurosci.* 17:6256-6263).

10

15

20

25

30

The signal peptide prediction program SIGNALP (Nielsen et al. (1997) *Protein Engineering* 10:1-6) predicted that human TANGO 265 protein includes a 31 amino acid signal peptide (amino acid residues 1 to 31 of SEQ ID NO: 19; SEQ ID NO: 20) preceding the mature TANGO 265 protein (amino acid residues 32 to 761 of SEQ ID NO: 19; SEQ ID NO: 21). Human TANGO 265 protein includes an extracellular domain (amino acid residues 32 to 683 of SEQ ID NO: 19; SEQ ID NO: 22); a transmembrane domain (amino acid residues 684 to 704 of SEQ ID NO: 19; SEQ ID NO: 23); and a cytoplasmic domain (amino acid residues 705 to 761 of SEQ ID NO: 19; SEQ ID NO: 24).

Figure 3U depicts a hydrophilicity plot of human TANGO 265 protein. Relatively hydrophobic regions are above the dashed horizontal line, and relatively hydrophilic regions are below the dashed horizontal line. The hydrophobic region which corresponds to amino acid residues 1 to 31 of SEQ ID NO: 19 is the signal sequence of human TANGO 265 (SEQ ID NO: 20). The hydrophobic region which corresponds to amino acid residues 684 to 704 of SEQ ID NO: 19 is the transmembrane domain of human TANGO 265 (SEQ ID NO: 23). As described elsewhere herein, relatively hydrophilic regions are generally located at or near the surface of a protein, and are more frequently effective immunogenic epitopes than are relatively hydrophobic regions. For example, the region of human TANGO 265 protein from about amino acid residue 350 to about amino acid

residue 375 appears to be located at or near the surface of the protein, while the region from about amino acid residue 230 to about amino acid residue 250 appears not to be located at or near the surface.

The predicted molecular weight of human TANGO 265 protein without modification and prior to cleavage of the signal sequence is about 83.6 kilodaltons. The predicted molecular weight of the mature human TANGO 265 protein without modification and after cleavage of the signal sequence is about 80.2 kilodaltons.

5

10

15

20

25

30

Chromosomal mapping was performed by computerized comparison of TANGO 265 cDNA sequences against a chromosomal mapping database in order to identify the approximate location of the gene encoding human TANGO 265 protein. This analysis indicated that the gene was located on chromosome 1 between markers D1S305 and D1S2635.

Biological function of TANGO 265 proteins, nucleic acids, and modulators thereof

TANGO 265 proteins are involved in disorders which affect both tissues in which they are normally expressed and tissues in which they are normally not expressed. Based on the observation that TANGO 265 is expressed in human fetal spleen, involvement of TANGO 202 protein in immune system development and modulation is indicated.

The presence of the sema domain in TANGO 265 indicates that this protein is involved in development of neuronal and epithelial tissues and also functions as a repellant protein which guides axonal development. TANGO 265 modulates nerve growth and regeneration and also modulates growth and regeneration of other epithelial tissues.

The observation that TANGO 265 shares significant identity with murine semaphorin B suggests that it has activity identical or analogous to the activity of this protein. These observations indicate that TANGO 265 modulates growth, proliferation, survival, differentiation, and activity of neuronal cells and immune system cells. Thus, TANGO 265 protein is useful, for example, for guiding neural axon development, for modulating differentiation of cells of the

immune system, for modulating cytokine production by cells of the immune system, for modulating reactivity of cells of the immune system toward cytokines, for modulating initiation and persistence of an inflammatory response, and for modulating proliferation of epithelial cells.

5

10

15

20

25

30

TANGO 273

A cDNA clone (designated jthoc028g06) encoding at least a portion of human TANGO 273 protein was isolated from a lipopolysaccharide- (LPS-)stimulated human osteoblast cDNA library. The corresponding murine cDNA clone (designated jtmoa001c04) was isolated from an LPS-stimulated murine osteoblast cDNA library. The human and murine TANGO 273 proteins are predicted by structural analysis to be transmembrane proteins.

The full length of the cDNA encoding human TANGO 273 protein (Figure 4; SEQ ID NO: 25) is 2964 nucleotide residues. The ORF of this cDNA, nucleotide residues 135 to 650 of SEQ ID NO: 25 (i.e., SEQ ID NO: 26), encodes a 172-amino acid transmembrane protein (Figure 4; SEQ ID NO: 27).

The invention thus includes purified human TANGO 273 protein, both in the form of the immature 172 amino acid residue protein (SEQ ID NO: 27) and in the form of the mature 150 amino acid residue protein (SEQ ID NO: 29). The invention also includes purified murine TANGO 273 protein, both in the form of the immature 172 amino acid residue protein (SEQ ID NO: 74) and in the form of the mature 150 amino acid residue protein (SEQ ID NO: 44). Mature human or murine TANGO 273 proteins can be synthesized without the signal sequence polypeptide at the amino terminus thereof, or they can be synthesized by generating immature TANGO 273 protein and cleaving the signal sequence therefrom.

In addition to full length mature and immature human and murine TANGO 273 proteins, the invention includes fragments, derivatives, and variants of these TANGO 273 proteins, as described herein. These proteins, fragments, derivatives, and variants are collectively referred to herein as polypeptides of the invention or proteins of the invention.

The invention also includes nucleic acid molecules which encode a polypeptide of the invention. Such nucleic acids include, for example, a DNA

molecule having the nucleotide sequence listed in SEQ ID NO: 25 or some portion thereof or SEQ ID NO: 73 or some portion thereof, such as the portion which encodes mature TANGO 273 protein, immature TANGO 273 protein, or a domain of TANGO 273 protein. These nucleic acids are collectively referred to as nucleic acids of the invention.

5

10

15

20

25

30.

TANGO 273 proteins and nucleic acid molecules encoding them comprise a family of molecules having certain conserved structural and functional features. This family includes, by way of example, the human and murine TANGO 273 proteins.

A common domain of TANGO 273 proteins is a signal sequence. As used herein, a signal sequence includes a peptide of at least about 10 amino acid residues in length which occurs at the amino terminus of membrane-bound proteins and which contains at least about 45% hydrophobic amino acid residues such as alanine, leucine, isoleucine, phenylalanine, proline, tyrosine, tryptophan, or valine. In a preferred embodiment, a signal sequence contains at least about 10 to 35 amino acid residues, preferably about 10 to 20 amino acid residues, and has at least about 35-60%, more preferably 40-50%, and more preferably at least about 45% hydrophobic residues. A signal sequence serves to direct a protein containing such a sequence to a lipid bilayer. Thus, in one embodiment, a TANGO 273 protein contains a signal sequence corresponding to amino acid residues 1 to 22 of SEQ ID NO: 27 (SEQ ID NO: 28) or to amino acid residues 1 to 22 of SEQ ID NO: 74. The signal sequence is cleaved during processing of the mature protein.

TANGO 273 proteins can also include an extracellular domain. The human TANGO 273 protein extracellular domain is located from about amino acid residue 23 to about amino acid residue 60 of SEQ ID NO: 27, and the murine TANGO 273 protein extracellular domain is located from about amino acid residue 23 to about amino acid residue 60 of SEQ ID NO: 74.

The present invention also includes TANGO 273 proteins having a transmembrane domain. As used herein, a "transmembrane domain" refers to an amino acid sequence having at least about 15 to 30 amino acid residues in length and which contains at least about 65-70% hydrophobic amino acid residues such as alanine, leucine, phenylalanine, protein, tyrosine, tryptophan, or valine. In a

preferred embodiment, a transmembrane domain contains at least about 15 to 20 amino acid residues, preferably about 20 to 25 amino acid residues, and has at least about 60-80%, more preferably 65-75%, and more preferably at least about 70% hydrophobic residues. Thus, in one embodiment, a human TANGO 273 protein of the invention contains a transmembrane domain corresponding to about amino acid residues 61 to 81 of SEQ ID NO: 27 (SEQ ID NO: 31). In another embodiment, a murine TANGO 273 protein of the invention contains a transmembrane domain corresponding to about amino acid residues 61 to 81 of SEQ ID NO: 74.

5

10

15

20

In addition, TANGO 273 proteins include a cytoplasmic domain. The human TANGO 273 cytoplasmic domain is located from about amino acid residue 82 to amino acid residue 172 of SEQ ID NO: 27 (SEQ ID NO: 32), and the murine TANGO 273 cytoplasmic domain is located from about amino acid residue 82 to amino acid residue 172 of SEQ ID NO: 74.

TANGO 273 proteins typically comprise a variety of potential post-translational modification sites (often within an extracellular domain), such as those described herein in Tables VII and VIII, as predicted by computerized sequence analysis of human and murine TANGO 273 proteins using amino acid sequence comparison software (comparing the amino acid sequence of TANGO 273 with the information in the PROSITE database {rel. 12.2; Feb, 1995} and the Hidden Markov Models database {Rel. PFAM 3.3}). In certain embodiments, a protein of the invention has at least 1, 2, 3, 4, 5, or all 6 of the post-translational modification sites listed in Table VII. In other embodiments, the protein of the invention has at least 1, 2, 3, 4, 5, 6, or all 7 of the post-translational modification sites listed in Table VIII.

Table VII

| Type of Potential Modification Site | Amino Acid Residues of | Amino Acid |
|---------------------------------------|------------------------|------------|
| · or Domain | SEQ ID NO: 27 | Sequence |
| N-glycosylation site | 97 to 100 | NVSY |
| Casein kinase II phosphorylation site | 41 to 44 | SYED |
| N-myristoylation site | 31 to 36 | GLYPTY |
| | 47 to 52 | GSRCCV |
| | 70 to 75 | GVLFCC |
| * y . | 131 to 136 | GNSMAM |
| Src Homology 3 (SH3) domain binding | 86 to 90 | YPPPL |
| site | 103 to 107 | QPPNP |
| · | 113 to 117 | QPGPP |
| | 121 to 125 | DPGGP |
| | 1,40 to 145 | VPPNSP |
| | 151 to 155 | СРРРР |
| | 160 to 164 | ТРРРР |

Table VIII

| Type of Potential Modification Site or Domain | Amino Acid Residues of SEQ ID NO: 74 | Amino Acid Sequence |
|---|---|------------------------|
| N-glycosylation site | 97 to 100 | NVSY |
| Casein kinase II phosphorylation site | 41 to 44 | SYED |
| N-myristoylation site | 31 to 36 | GLYPTY |
| | 47 to 52 | GSRCCV |
| | 70 to 75 | GVLFCC |
| | 131 to 136 | GNTMAM |

Table VIII (Continued)

| Src Homology 3 (SH3) domain binding | 86 to 90 | YPPPL |
|-------------------------------------|------------|-------|
| site | 103 to 107 | QPPNP |
| | 115 to 119 | GPPYY |
| | 121 to 125 | DPGGP |
| | 141 to 145 | QPNSP |
| | 151 to 155 | YPPPP |
| | 160 to 164 | ТРРРР |
| Amidation site | 1 to 4 | MGRR |

The amino acid sequence of TANGO 273 protein includes about

seven potential proline-rich Src homology 3 (SH3) domain binding sites nearer the
cytoplasmic portion of the protein. SH3 domains mediate specific assembly of
protein complexes, presumably by interacting with proline-rich protein domains
(Morton and Campbell (1994) Curr. Biol. 4:615-617). SH3 domains also mediate
interactions between proteins involved in transmembrane signal transduction.

Coupling of proteins mediated by SH3 domains has been implicated in a variety of

Coupling of proteins mediated by SH3 domains has been implicated in a variety of physiological systems, including those involving regulation of cell growth and proliferation, endocytosis, and activation of respiratory burst.

15

20

SH3 domains have been described in the art (e.g., Mayer et al. (1988) Nature 332:272-275; Musacchio et al. (1992) FEBS Lett. 307:55-61; Pawson and Schlessinger (1993) Curr. Biol. 3:434-442; Mayer and Baltimore (1993) Trends Cell Biol. 3:8-13; Pawson (1993) Nature 373:573-580), and occur in a variety of cytoplasmic proteins, including several (e.g., protein tyrosine kinases) involved in transmembrane signal transduction. Among the proteins in which one or more SH3 domains occur are protein tyrosine kinases such as those of the Src, Abl, Bkt, Csk and ZAP70 families, mammalian phosphatidylinositol-specific phospholipases C-gamma-1 and -2, mammalian phosphatidylinositol 3-kinase regulatory p85 subunit, mammalian Ras GTPase-activating protein (GAP), proteins which mediate binding of guanine nucleotide exchange factors and growth factor receptors (e.g., vertebrate

GRB2, Caenorhabditis elegans sem-5, and Drosophila DRK proteins), mammalian Vav oncoprotein, guanidine nucleotide releasing factors of the CDC 25 family (e.g., yeast CDC25, yeast SCD25, and fission yeast ste6 proteins), MAGUK proteins (e.g., mammalian tight junction protein ZO-1, vertebrate erythrocyte membrane protein p55, C. elegans protein lin-2, rat protein CASK, and mammalian synaptic proteins SAP90/PSD-95, CHAPSYN-110/PSD-93, SAP97/DLG1, and SAP102), proteins which interact with vertebrate receptor protein tyrosine kinases (e.g., mammalian cytoplasmic protein Nck and oncoprotein Crk), chicken Src substrate p80/85 protein (cortactin), human hemopoietic lineage cell specific protein Hs1, mammalian dihydrouridine-sensitive L-type calcium channel beta subunit, human myasthenic syndrome antigen B (MSYB), mammalian neutrophil cytosolic activators of NADPH oxidase (e.g., p47 {NCF-1}, p67 {NCF-2}, and C. elegans protein B0303.7) myosin heavy chains (MYO3) from amoebae, from slime molds, and from yeast, vertebrate and *Drosophila* spectrin and fodrin alpha chain proteins, human amphiphysin, yeast actin-binding proteins ABP1 and SLA3, yeast protein BEM1, fission yeast protein scd2 (ral3), yeast BEM1-binding proteins BOI2 (BEB1) and BOB1 (BOI1), yeast fusion protein FUS1, yeast protein RSV167, yeast protein SSU81, yeast hypothetical proteins YAR014c, YFR024c, YHL002w, YHR016c, YJL020C, and YHR114w, hypothetical fission yeast protein SpAC12C2.05c, and C. elegans hypothetical protein F42H10.3. Of these proteins, multiple SH3 domains occur in vertebrate GRB2 protein, C. elegans sem-5 protein, Drosophila DRK protein, oncoprotein Crk, mammalian neutrophil cytosolic activators of NADPH oxidase p47 and p67, yeast protein BEM1, fission yeast protein scd2, yeast hypothetical protein YHR114w, mammalian cytoplasmic protein Nck, C. elegans neutrophil cytosolic activator of NADPH oxidase B0303.7, and yeast actin-binding protein SLA1. Of these proteins, three or more SH3 domains occur in mammalian cytoplasmic protein Nck, C. elegans neutrophil cytosolic activator of NADPH oxidase B0303.7, and yeast actin-binding protein SLA1. The presence of SH3 domain binding sites in TANGO 273 indicates that TANGO 273 interacts with one or more of these and other SH3 domain-containing proteins and 30 is thus involved in physiological processes in which one or more of these or other SH3 domain-containing proteins are involved.

10

15

20

25

The signal peptide prediction program SIGNALP (Nielsen et al. (1997) Protein Engineering 10:1-6) predicted that human TANGO 273 protein includes a 22 amino acid signal peptide (amino acid residues 1 to 22 of SEQ ID NO: 27; SEQ ID NO: 28) preceding the mature TANGO 273 protein (amino acid residues 23 to 172 of SEQ ID NO: 27; SEQ ID NO: 29). Human TANGO 273 protein includes an extracellular domain (amino acid residues 23 to 60 of SEQ ID NO: 27; SEQ ID NO: 30); a transmembrane domain (amino acid residues 61 to 81 of SEQ ID NO: 27; SEQ ID NO: 31); and a cytoplasmic domain (amino acid residues 82 to 172 of SEQ ID NO: 27; SEQ ID NO: 32).

5

10

15

20

25

protein. Relatively hydrophobic regions are above the dashed horizontal line, and relatively hydrophobic regions are below the dashed horizontal line. The hydrophobic region which corresponds to amino acid residues 1 to 22 of SEQ ID NO: 27 is the signal sequence of human TANGO 273 (SEQ ID NO: 28). The hydrophobic region which corresponds to amino acid residues 61 to 81 of SEQ ID NO: 27 is the transmembrane domain of human TANGO 273 (SEQ ID NO: 31). As described elsewhere herein, relatively hydrophilic regions are generally located at or near the surface of a protein, and are more frequently effective immunogenic epitopes than are relatively hydrophobic regions. For example, the region of human TANGO 273 protein from about amino acid residue 100 to about amino acid residue 120 appears to be located at or near the surface of the protein, while the region from about amino acid residue 130 to about amino acid residue 140 appears not to be located at or near the surface.

Chromosomal mapping was performed by computerized comparison of TANGO 273 cDNA sequences against a chromosomal mapping database in order to identify the approximate location of the gene encoding human TANGO 273 protein. This analysis indicated that the gene was located on chromosome 7 between markers D7S2467 and D7S2552.

The predicted molecular weight of human TANGO 273 protein
without modification and prior to cleavage of the signal sequence is about 19.2
kilodaltons. The predicted molecular weight of the mature human TANGO 273

protein without modification and after cleavage of the signal sequence is about 16.8 kilodaltons.

Northern analysis experiments indicated that mRNA corresponding to the cDNA encoding TANGO 273 is expressed in the tissues listed in Table VIIa, wherein "++" indicates moderate expression and "+" indicates lower expression.

5

10

15

20

Table VIIa

| Animal | Tissue | Relative Level of Expression |
|--------|-----------------|------------------------------|
| Human | heart | ++ |
| | brain | ++ |
| | skeletal muscle | ++ |
| | pancreas | ++ |
| | placenta | . + |
| | lung | + |
| | liver | + |
| | kidney | + |

The full length of the cDNA encoding murine TANGO 273 protein (Figure 4; SEQ ID NO: 72) is 2915 nucleotide residues. The ORF of this cDNA, nucleotide residues 137 to 650 of SEQ ID NO: 72 (i.e., SEQ ID NO: 73), encodes a 172-amino acid transmembrane protein (Figure 4; SEQ ID NO: 74).

The signal peptide prediction program SIGNALP (Nielsen et al. (1997) Protein Engineering 10:1-6) predicted that murine TANGO 273 protein includes a 22 amino acid signal peptide (amino acid residues 1 to 22 of SEQ ID NO: 74) preceding the mature TANGO 273 protein (amino acid residues 23 to 172 of SEQ ID NO: 74; SEQ ID NO: 44). Murine TANGO 273 protein includes an extracellular domain (amino acid residues 23 to 60 of SEQ ID NO: 74); a transmembrane domain (amino acid residues 61 to 81 of SEQ ID NO: 74); and a cytoplasmic domain (amino acid residues 82 to 172 of SEQ ID NO: 74).

Figure 4J depicts a hydrophilicity plot of murine TANGO 273 protein. Relatively hydrophobic regions are above the dashed horizontal line, and

relatively hydrophilic regions are below the dashed horizontal line. The hydrophobic region which corresponds to amino acid residues 1 to 22 of SEQ ID NO: 74 is the signal sequence of murine TANGO 273. As described elsewhere herein, relatively hydrophilic regions are generally located at or near the surface of a protein, and are more frequently effective immunogenic epitopes than are relatively hydrophobic regions. For example, the region of murine TANGO 273 protein from about amino acid residue 100 to about amino acid residue 120 appears to be located at or near the surface of the protein, while the region from about amino acid residue 130 to about amino acid residue 140 appears not to be located at or near the surface.

5

10

15

20

25

30

The predicted molecular weight of murine TANGO 273 protein without modification and prior to cleavage of the signal sequence is about 19.4 kilodaltons. The predicted molecular weight of the mature murine TANGO 273 protein without modification and after cleavage of the signal sequence is about 17.1 kilodaltons.

In situ analysis of murine TANGO 273 mRNA indicated that TANGO 273 is expressed with central nervous system (CNS) tissues during embryogenesis and into adulthood. Expression of TANGO 273 is widely observed in murine CNS tissues, including brain, spinal cord, eye, and olfactory epithelium at all embryonic ages examined (i.e., at embryonic days 13.5, 14.5, 15.5, 16.5, and 18.5 and at post-natal day 1.5).

Human and murine TANGO 273 cDNA sequences exhibit significant nucleotide sequence identity with an expressed sequence tag (EST) isolated from a library of ESTs corresponding to proteins secreted from prostate tissue, as described in PCT publication number WO 99/06550, published February 11, 1999.

Human and murine TANGO 273 proteins exhibit considerable sequence similarity, as indicated herein in Figure 4H. Figure 4H depicts an alignment of human and murine TANGO 273 protein amino acid sequences (SEQ ID NOs: 27 and 74, respectively). In this alignment (pam120.mat scoring matrix, gap penalties

-12/-4), the proteins are 89.5% identical. Alignment of the ORF encoding human TANGO 273 protein and the ORF encoding murine TANGO 273 protein using the same software and parameters indicated that the nucleotide sequences are 84.1% identical.

5

10

15

20

25

30

Biological function of TANGO 273 proteins, nucleic acids, and modulators thereof

cDNAs encoding the human and murine TANGO 273 proteins were each isolated from LPS-stimulated osteoblast cDNA libraries. These proteins are involved in bone-related metabolism, homeostasis, and development disorders. Thus, proteins and nucleic acids of the invention which are identical to, similar to, or derived from human and murine TANGO 273 proteins and nucleic acids encoding them are useful for preventing, diagnosing, and treating, among others, bone-related disorders such as osteoporosis, cancer, skeletal development disorders, bone fragility, and the like.

Expression of TANGO 273 in heart, brain, skeletal muscle, and pancreas, placenta, lung, liver, and kidney tissues is an indication that TANGO 273 proteins, nucleic acids encoding them, and agents that modulate activity or expression of either of these can be used to modulate growth, proliferation, survival, differentiation, adhesion, and activity of cells of these tissues, or to prognosticate, diagnose, and treat one or more disorders which affect these tissues.

The fact that TANGO 273 is expressed at high levels in neurological tissues is an indication that TANGO 273 proteins, nucleic acids, and modulators thereof can be used to modulate proliferation, differentiation, or function of neurological cells in these tissues (e.g., neuronal cells). Thus, TANGO 273 proteins, nucleic acids, and modulators thereof can be used to prognosticate, diagnose, and treat one or more neurological disorders. Examples of such disorders include CNS disorders, CNS-related disorders, focal brain disorders, global-diffuse cerebral disorders, and other neurological and cerebrovascular disorders.

CNS disorders include, but are not limited to cognitive and neurodegenerative disorders such as Alzheimer's disease, senile dementia,

Huntington's disease, amyotrophic lateral sclerosis, and Parkinson's disease, as well as Gilles de la Tourette's syndrome, autonomic function disorders such as hypertension and sleep disorders (e.g., insomnia, hypersomnia, parasomnia, and sleep apnea); neuropsychiatric disorders (e.g., schizophrenia, schizoaffective disorder, attention deficit disorder, dysthymic disorder, major depressive disorder, mania, and obsessive-compulsive disorder); psychoactive substance use disorders; anxiety; panic disorder; and bipolar affective disorders (e.g., severe bipolar affective disorder and bipolar affective disorder with hypomania and major depression).

5

10

15

20

25

30

CNS-related disorders include disorders associated with developmental, cognitive, and autonomic neural and neurological processes, such as pain, appetite, long term memory, and short term memory.

Exemplary focal brain disorders include aphasia, apraxia, agnosia, and amnesias (e.g., posttraumatic amnesia, transient global amnesia, and psychogenic amnesia). Global-diffuse cerebral disorders with which TANGO 273 can be associated include coma, stupor, obtundation, and disorders of the reticular formation.

Other neurological disorders with which TANGO 273 can be associated include ischemic syndromes (e.g., stroke), hypertensive encephalopathy, hemorrhagic disorders, and disorders involving aberrant function of the blood-brain barrier (e.g., CNS infections such as meningitis and encephalitis, aseptic meningitis, metastasis of non-CNS tumor cells into the CNS, various pain disorders such as migraine, blindness and other vision problems, and CNS-related adverse drug reactions such as head pain, sleepiness, and confusion). TANGO 273 proteins, nucleic acids encoding them, and agents that modulate activity or expression of either of these can be used to prognosticate, diagnose, and treat one or more of these disorders.

Developmental regulation of TANGO 273 expression in fetal neurological tissues, as described herein, is an indication that TANGO 273 proteins, nucleic acids, and modulators thereof can be used to prognosticate, diagnose, and treat one or more disorders which involve aberrant fetal neurological development. Examples of such disorders include blindness,

deafness, fetal death, mental retardation, dysraphia, anencephaly, malformation of cerebral hemispheres, encephalocele, porencephaly, hydranencephaly, hydrocephalus, and spina bifida.

The fact that TANGO 273 is expressed in tissues which were exposed to LPS indicates that TANGO 273 mediates one or more physiological responses of cells to bacterial infection. Thus, TANGO 273 is involved in one or more of detection of bacteria in a tissue in which it is expressed, movement of cells with relation to sites of bacterial infection, production of biological molecules which inhibit bacterial infection, and production of biological molecules which alleviate cellular or other physiological damage wrought by bacterial infection.

5

10

15

20

25

30

Presence in TANGO 273 protein of multiple SH3 domain binding sites indicates that TANGO 273 protein interacts with one or more SH3 domain-containing proteins. Thus, TANGO 273 protein mediates binding of proteins (i.e., binding of proteins to TANGO 273 and to one another to form protein complexes) in cells in which it is expressed. TANGO 273 is also involved in transduction of signals between the exterior environment of cells (i.e., including from other cells) and the interior of cells in which it is expressed. TANGO 273 mediates regulation of cell growth and proliferation, endocytosis, activation of respiratory burst, and other physiological processes triggered by transmission of a signal via a protein with which TANGO 273 interacts.

Sequence similarity of TANGO 273 cDNA with an EST expressed in prostate tissue indicates that TANGO 273 can be expressed in prostate tissue, and can thus be involved in disorders of the prostate. Thus, TANGO 273 proteins, nucleic acids encoding them, and agents that modulate activity or expression of either of these can be used to treat prostate disorders. Examples of prostate disorders which can be treated in this manner include inflammatory prostatic diseases (e.g., acute and chronic prostatitis and granulomatous prostatitis), prostatic hyperplasia (e.g., benign prostatic hypertrophy or hyperplasia), and prostate tumors (e.g., carcinomas).

In another example, TANGO 273 polypeptides, nucleic acids, or modulators thereof, can be used to treat cardiovascular disorders, such as ischemic heart disease (e.g., angina pectoris, myocardial infarction, and chronic ischemic

heart disease), hypertensive heart disease, pulmonary heart disease, valvular heart disease (e.g., rheumatic fever and rheumatic heart disease, endocarditis, mitral valve prolapse, and aortic valve stenosis), congenital heart disease (e.g., valvular and vascular obstructive lesions, atrial or ventricular septal defect, and patent ductus arteriosus), or myocardial disease (e.g., myocarditis, congestive cardiomyopathy, and hypertrophic cardiomyopathy).

5

10

15

20

25

30

In another example, TANGO 273 polypeptides, nucleic acids, or modulators thereof, can be used to treat disorders of the brain, such as cerebral edema, hydrocephalus, brain herniations, iatrogenic disease (due to, e.g., infection, toxins, or drugs), inflammations (e.g., bacterial and viral meningitis, encephalitis, and cerebral toxoplasmosis), cerebrovascular diseases (e.g., hypoxia, ischemia, and infarction, intracranial hemorrhage and vascular malformations, and hypertensive encephalopathy), and tumors (e.g., neuroglial tumors, neuronal tumors, tumors of pineal cells, meningeal tumors, primary and secondary lymphomas, intracranial tumors, and medulloblastoma), and to treat injury or trauma to the brain.

In another example, TANGO 273 polypeptides, nucleic acids, or modulators thereof, can be used to treat disorders of skeletal muscle, such as muscular dystrophy (e.g., Duchenne muscular dystrophy, Becker muscular dystrophy, Emery-Dreifuss muscular dystrophy, limb-girdle muscular dystrophy, facioscapulohumeral muscular dystrophy, myotonic dystrophy, oculopharyngeal muscular dystrophy, distal muscular dystrophy, and congenital muscular dystrophy), motor neuron diseases (e.g., amyotrophic lateral sclerosis, infantile progressive spinal muscular atrophy, intermediate spinal muscular atrophy, spinal bulbar muscular atrophy, and adult spinal muscular atrophy), myopathies (e.g., inflammatory myopathies such as dermatomyositis and polymyositis, myotonia congenita, paramyotonia congenita, central core disease, nemaline myopathy, myotubular myopathy, and periodic paralysis), and metabolic diseases of muscle (e.g., phosphorylase deficiency, acid maltase deficiency, phosphofructokinase deficiency, debrancher enzyme deficiency, mitochondrial myopathy, carnitine deficiency, carnitine palmityl transferase deficiency, phosphoglycerate kinase deficiency, phosphoglycerate mutase deficiency, lactate dehydrogenase deficiency, and myoadenylate deaminase deficiency).

In another example, TANGO 273 polypeptides, nucleic acids, or modulators thereof, can be used to treat pancreatic disorders, such as pancreatitis (e.g., acute hemorrhagic pancreatitis and chronic pancreatitis), pancreatic cysts (e.g., congenital cysts, pseudocysts, and benign or malignant neoplastic cysts), pancreatic tumors (e.g., pancreatic carcinoma and adenoma), diabetes mellitus (e.g., insulin- and non-insulin-dependent types, impaired glucose tolerance, and gestational diabetes), or islet cell tumors (e.g., insulinomas, adenomas, Zollinger-Ellison syndrome, glucagonomas, and somatostatinoma).

5

10

15

20

In another example, TANGO 273 polypeptides, nucleic acids, or modulators thereof, can be used to treat placental disorders, such as toxemia of pregnancy (e.g., preeclampsia and eclampsia), placentitis, or spontaneous abortion.

In another example, TANGO 273 polypeptides, nucleic acids, or modulators thereof, can be used to treat pulmonary disorders, such as atelectasis, cystic fibrosis, rheumatoid lung disease, pulmonary congestion or edema, chronic obstructive airway disease (e.g., emphysema, chronic bronchitis, bronchial asthma, and bronchiectasis), diffuse interstitial diseases (e.g., sarcoidosis, pneumoconiosis, hypersensitivity pneumonitis, Goodpasture's syndrome, idiopathic pulmonary hemosiderosis, pulmonary alveolar proteinosis, desquamative interstitial pneumonitis, chronic interstitial pneumonia, fibrosing alveolitis, hamman-rich syndrome, pulmonary eosinophilia, diffuse interstitial fibrosis, Wegener's granulomatosis, lymphomatoid granulomatosis, and lipid pneumonia), or tumors (e.g., bronchogenic carcinoma, bronchioalveolar carcinoma, bronchial carcinoid, hamartoma, and mesenchymal tumors).

In another example, TANGO 273 polypeptides, nucleic acids, or modulators thereof, can be used to treat hepatic (liver) disorders, such as jaundice, hepatic failure, hereditary hyperbilirubinemias (e.g., Gilbert's syndrome, Crigler-Naijar syndromes, and Dubin-Johnson and Rotor's syndromes), hepatic circulatory disorders (e.g., hepatic vein thrombosis and portal vein obstruction and thrombosis) hepatitis (e.g., chronic active hepatitis, acute viral hepatitis, and toxic and druginduced hepatitis) cirrhosis (e.g., alcoholic cirrhosis, biliary cirrhosis, and hemochromatosis), or malignant tumors (e.g., primary carcinoma, hepatoblastoma, and angiosarcoma).

In another example, TANGO 273 polypeptides, nucleic acids, or modulators thereof, can be used to treat renal (kidney) disorders, such as glomerular diseases (e.g., acute and chronic glomerulonephritis, rapidly progressive glomerulonephritis, nephrotic syndrome, focal proliferative glomerulonephritis, glomerular lesions associated with systemic disease such as 5 systemic lupus erythematosus, Goodpasture's syndrome, multiple myeloma, diabetes, neoplasia, sickle cell disease, and chronic inflammatory diseases), tubular diseases (e.g., acute tubular necrosis and acute renal failure, polycystic renal disease, medullary sponge kidney, medullary cystic disease, nephrogenic diabetes, and renal tubular acidosis), tubulointerstitial diseases (e.g., 10 pyelonephritis, drug and toxin induced tubulointerstitial nephritis, hypercalcemic nephropathy, and hypokalemic nephropathy) acute and rapidly progressive renal failure, chronic renal failure, nephrolithiasis, vascular diseases (e.g., hypertension and nephrosclerosis, microangiopathic hemolytic anemia, atheroembolic renal disease, diffuse cortical necrosis, and renal infarcts), or tumors (e.g., renal cell 15 carcinoma and nephroblastoma).

TANGO 286

20

25

30

A cDNA clone (designated jthkf042e03) encoding at least a portion of human TANGO 286 protein was isolated from a human keratinocyte cDNA library. The human TANGO 286 protein is predicted by structural analysis to be a secreted protein.

The full length of the cDNA encoding TANGO 286 protein (Figure 5; SEQ ID NO: 33) is 1980 nucleotide residues. The ORF of this cDNA, nucleotide residues 133 to 1497 of SEQ ID NO: 33 (i.e., SEQ ID NO: 34), encodes a 455-amino acid secreted protein (Figure 5; SEQ ID NO: 35).

The invention thus includes purified TANGO 286 protein, both in the form of the immature 455 amino acid residue protein (SEQ ID NO: 35) and in the form of the mature 432 amino acid residue protein (SEQ ID NO: 37). Mature TANGO 286 protein can be synthesized without the signal sequence polypeptide at the amino terminus thereof, or it can be synthesized by generating immature TANGO 286 protein and cleaving the signal sequence therefrom.

In addition to full length mature and immature TANGO 286 proteins, the invention includes fragments, derivatives, and variants of these TANGO 286 proteins, as described herein. These proteins, fragments, derivatives, and variants are collectively referred to herein as polypeptides of the invention or proteins of the invention.

5

10

15

20

25

30

The invention also includes nucleic acid molecules which encode a polypeptide of the invention. Such nucleic acids include, for example, a DNA molecule having the nucleotide sequence listed in SEQ ID NO: 33 or some portion thereof, such as the portion which encodes mature TANGO 286 protein, immature TANGO 286 protein, or a domain of TANGO 286 protein. These nucleic acids are collectively referred to as nucleic acids of the invention.

TANGO 286 proteins and nucleic acid molecules encoding them comprise a family of molecules having certain conserved structural and functional features.

A common domain of TANGO 286 proteins is a signal sequence. As used herein, a signal sequence includes a peptide of at least about 10 amino acid residues in length which occurs at the amino terminus of membrane-bound proteins and which contains at least about 45% hydrophobic amino acid residues such as alanine, leucine, isoleucine, phenylalanine, proline, tyrosine, tryptophan, or valine. In a preferred embodiment, a signal sequence contains at least about 10 to 35 amino acid residues, preferably about 10 to 20 amino acid residues, and has at least about 35-60%, more preferably 40-50%, and more preferably at least about 45% hydrophobic residues. A signal sequence serves to direct a protein containing such a sequence to a lipid bilayer. Thus, in one embodiment, a TANGO 286 protein contains a signal sequence corresponding to amino acid residues 1 to 23 of SEQ ID NO: 35 (SEQ ID NO: 36). The signal sequence is cleaved during processing of the mature protein.

TANGO 286 is a secreted soluble protein (i.e., a secreted protein having a single extracellular domain), as indicated by computerized sequence analysis and comparison of the amino acid sequence of TANGO 286 with related proteins, such as the soluble proteins designated bactericidal permeability

increasing (BPI) protein and recombinant endotoxin neutralizing polypeptide (RENP).

5

10

TANGO 286 proteins typically comprise a variety of potential post-translational modification sites (often within an extracellular domain), such as those described herein in Table IX, as predicted by computerized sequence analysis of TANGO 286 proteins using amino acid sequence comparison software (comparing the amino acid sequence of TANGO 286 with the information in the PROSITE database {rel. 12.2; Feb, 1995} and the Hidden Markov Models database {Rel. PFAM 3.3}). In certain embodiments, a protein of the invention has at least 1, 2, 4, 6, 10, 15, or 20 or more of the post-translational modification sites listed in Table IX.

Table IX

| Type of Potential Modification Site | Amino Acid Residues of | Amino Acid |
|---------------------------------------|------------------------|------------|
| or Domain | SEQ ID NO: 35 | Sequence |
| N-glycosylation site | 79 to 82 | NFSN |
| | 92 to 95 | NTSL |
| | 113 to 116 | NIST |
| | 161 to 164 | NLST |
| | 173 to 176 | NYTL |
| | 205 to 208 | NLTD |
| | 249 to 252 | NLTL |
| | 303 to 306 | NFTL |
| | 320 to 323 | NSTV |
| | 363 to 366 | NRSN |
| Protein kinase C phosphorylation site | 35 to 37 | TQR |
| | 362 to 364 | SNR |
| | 429 to 431 | SSK |

15

Table IX (Continued)

| Casein kinase II phosphorylation site | 63 to 66 | SGSE |
|---------------------------------------|------------|------------|
| - | 130 to 133 | SFAE |
| | 163 to 166 | STLE |
| | 169 to 172 | TKID |
| | 175 to 178 | TLLD |
| 7 | 183 to 186 | SSPE |
| | 253 to 256 | STEE |
| | 321 to 324 | STVE |
| * | 365 to 368 | SNIE |
| | 409 to 412 | SDIE |
| N-myristoylation site | 42 to 47 | GVQAGM |
| | 269 to 274 | GNVLSR |
| Lipid-binding serum glycoprotein | 12 to 427 | see Fig. 5 |
| domain | | |

Certain lipid-binding serum glycoproteins, such as LPS-binding

protein (LBP), bactericidal permeability-increasing protein (BPI), cholesteryl ester transfer protein (CETP), and phospholipid transfer protein (PLTP), share regions of sequence similarity which are herein designated a lipid-binding serum glycoprotein domain (Schumann et al., (1990) Science 249:1429-1431; Gray et al., (1989) J. Biol. Chem. 264:9505-9509; Day et al., (1994) J. Biol. Chem. 269:9388-9391). The consensus pattern of lipid-binding serum glycoprotein domains is as follows (using standard single letter amino acid abbreviations wherein X is any amino acid residue).

-(P or A)-(G or A)-(L or I or V or M or C)-
$$X_2$$
-R-(I or V)-(S or T)- X_3 -L- $X_{(4 \text{ or } 5)}$ -(E or Q)- X_4 -(L or I or V or M)- $X_{(0 \text{ or } 1)}$ -(E or Q or K)- X_8 -P-(e.g., amino acid residues 28-60 of SEQ ID NO: 35).

Proteins in which a lipid-binding serum glycoprotein domain occurs are often structurally related and exhibit related physiological activities. LBP binds to lipid A moieties of bacterial LPS and, once bound thereto, induces secretion of α-tumor necrosis factor, apparently by interacting with the CD14 receptor. BPI also

binds LPS and exerts a cytotoxic effect on Gram-negative bacteria (Elsbach, (1998) J. Leukoc. Biol. 64:14-18). CETP is involved in transfer of insoluble cholesteryl esters during reverse cholesterol transport. PLTP appears to be involved in phospholipid transport and modulation of serum HDL particles.

The signal peptide prediction program SIGNALP (Nielsen et al. (1997) *Protein Engineering* 10:1-6) predicted that TANGO 286 protein includes a 23 amino acid signal peptide (amino acid residues 1 to 23 of SEQ ID NO: 35; SEQ ID NO: 36) preceding the mature TANGO 286 protein (amino acid residues 24 to 455 of SEQ ID NO: 35; SEQ ID NO: 37). Human TANGO 286 protein is a secreted soluble protein.

5

10

15

20

25

30

Figure 5E depicts a hydrophilicity plot of TANGO 286 protein. Relatively hydrophobic regions are above the dashed horizontal line, and relatively hydrophilic regions are below the dashed horizontal line. As described elsewhere herein, relatively hydrophilic regions are generally located at or near the surface of a protein, and are more frequently effective immunogenic epitopes than are relatively hydrophobic regions. For example, the region of human TANGO 286 protein from about amino acid residue 420 to about amino acid residue 435 appears to be located at or near the surface of the protein, while the region from about amino acid residue 325 to about amino acid residue 345 appears not to be located at or near the surface.

The predicted molecular weight of TANGO 286 protein without modification and prior to cleavage of the signal sequence is about 50.9 kilodaltons. The predicted molecular weight of the mature TANGO 286 protein without modification and after cleavage of the signal sequence is about 48.2 kilodaltons.

The gene encoding human TANGO 286 protein was determined to be located on chromosome 22 by comparison of matching genomic clones such as the clones assigned GenBank Accession numbers W16806 and AL021937.

A portion of TANGO 286 protein exhibits significant amino acid homology with a region of the human chromosome region 22q12-13 genomic nucleotide sequence having GenBank Accession number AL021937. Alignment of a 45 kilobase nucleotide sequence encoding TANGO 286 with AL021937, however, indicated the presence in TANGO 286 of exons which differ from those

disclosed in L021937 (pam120.mat scoring matrix; gap penalties -12/-4). This region of chromosome 22 comprises an immunoglobulin lambda chain C (IGLC) pseudogene, the Ret finger protein-like 3 (RFPL3) and Ret finger protein-like 3 antisense (RFPL3S) genes, a gene encoding a novel immunoglobulin lambda chain

V family protein, a novel gene encoding a protein similar both to mouse RGDS protein (RALGDS, RALGEF, guanine nucleotide dissociation stimulator A) and to rabbit oncogene RSC, a novel gene encoding the human orthologue of worm F16A11.2 protein, a novel gene encoding a protein similar both to BPI and to rabbit liposaccharide-binding protein, and a 5'-portion of a novel gene. This region also comprises various ESTs, STSs, GSSs, genomic marker D22S1175, a ca repeat polymorphism and putative CpG islands. TANGO 286 protein thus shares one or more structural or functional features of these molecules.

15

20

25

30

TANGO 286 protein exhibits considerable sequence similarity with BPI protein, having 23.9% amino acid sequence identity therewith, as assessed using the ALIGN v. 2.0 computer software using a pam120.mat scoring matrix and gap penalties of -12/-4. TANGO 286 protein also exhibits considerable sequence similarity with recombinant endotoxin neutralizing polypeptide (RENP), having 24.5% amino acid sequence identity therewith, as assessed using the ALIGN software. Physiological activities of BPI protein and RENP have been described (e.g., Gabay et al., (1989) Proc. Natl. Acad. Sci. USA 86:5610-5614; Elsbach, (1998) J. Leukoc. Biol. 64:14-18; Mahadeva et al., (1997) Chest 112:1699-1701; International patent application WO96/34873). RENP, for example, binds LPS and neutralizes bacterial endotoxins. BPI, RENP, and other proteins in which a lipidbinding serum glycoprotein domain occurs bind LPS and neutralize bacterial endotoxins, and are therefore useful for preventing, detecting, and treating LPSrelated disorders such as shock, disseminated intravascular coagulation, anemia, thrombocytopenia, adult respiratory distress syndrome, renal failure, liver disease, and disorders associated with Gram negative bacterial infections. In addition to the physiological conditions described above, BPI protein is known to be involved in vasculitis and bronchiectasis, in that antibodies which bind specifically with BPI protein are present in at least some patients afflicted with these disorders (Mahadeva et al., supra).

Biological function of TANGO 286 proteins, nucleic acids, and modulators thereof

Expression of TANGO 286 in keratinocyte library indicates that this

protein is involved in a disorders which involve keratinocytes. Such disorders include, for example, disorders involving extracellular matrix abnormalities, dermatological disorders, ocular disorders, inappropriate hair growth (e.g., baldness), infections of the nails of the fingers and toes, scalp disorders (e.g., dandruff), and the like.

The fact that TANGO 286 protein contains a lipid-binding serum glycoprotein domain indicates that TANGO 286 is involved in one or more physiological processes in which these other lipid-binding serum glycoprotein domain-containing proteins are involved. Thus, TANGO 286 is involved in one or more of lipid transport, metabolism, serum lipid particle regulation, host antimicrobial defensive mechanisms, and the like.

Human TANGO 286 shares physiological functionality with other proteins in which a lipid-binding serum glycoprotein domains occurs (e.g., LBP, BPI protein, CETP, and PLTP). Based on the amino acid sequence similarity of TANGO 286 with BPI protein and with RENP, TANGO 286 protein exhibits physiological activities exhibited by these proteins. Thus, TANGO 286 proteins are useful for preventing, diagnosing, and treating, among others, lipid transport disorders, lipid metabolism disorders, disorders of serum lipid particle regulation, obesity, disorders involving insufficient or inappropriate host anti-microbial defensive mechanisms, vasculitis, bronchiectasis, LPS-related disorders such as shock, disseminated intravascular coagulation, anemia, thrombocytopenia, adult respiratory distress syndrome, renal failure, liver disease, and disorders associated with Gram negative bacterial infections, such as bacteremia, endotoxemia, sepsis, and the like.

30 TANGO 294

10

15

20

25

A cDNA clone (designated jthrc145g07) encoding at least a portion of human TANGO 294 protein was isolated from a human pulmonary artery

smooth muscle cell cDNA library. The human TANGO 294 protein is predicted by structural analysis to be a transmembrane membrane protein. No expression of DNA encoding TANGO 294 was detected in human heart, brain, placenta, lung, liver, skeletal muscle, kidney, or pancreas tissues.

The full length of the cDNA encoding TANGO 294 protein (Figure 6; SEQ ID NO: 45) is 2044 nucleotide residues. The ORF of this cDNA, nucleotide residues 126 to 1394 of SEQ ID NO: 45 (i.e., SEQ ID NO: 46), encodes a 423-amino acid transmembrane protein (Figure 6; SEQ ID NO: 47).

5

10

15

20

25

30

The invention includes purified TANGO 294 protein, both in the form of the immature 423 amino acid residue protein (SEQ ID NO: 47) and in the form of the mature 390 amino acid residue protein (SEQ ID NO: 49). Mature TANGO 294 protein can be synthesized without the signal sequence polypeptide at the amino terminus thereof, or it can be synthesized by generating immature TANGO 294 protein and cleaving the signal sequence therefrom.

In addition to full length mature and immature TANGO 294 proteins, the invention includes fragments, derivatives, and variants of TANGO 294 protein, as described herein. These proteins, fragments, derivatives, and variants are collectively referred to herein as polypeptides of the invention or proteins of the invention.

The invention also includes nucleic acid molecules which encode a polypeptide of the invention. Such nucleic acids include, for example, a DNA molecule having the nucleotide sequence listed in SEQ ID NO: 45 or some portion thereof, such as the portion which encodes mature TANGO 294 protein, immature TANGO 294 protein, or a domain of TANGO 294 protein. These nucleic acids are collectively referred to as nucleic acids of the invention.

TANGO 294 proteins and nucleic acid molecules encoding them comprise a family of molecules having certain conserved structural and functional features.

Also included within the scope of the invention are TANGO 294 proteins having a signal sequence. As used herein, a signal sequence includes a peptide of at least about 10 amino acid residues in length which occurs at the amino terminus of membrane-bound proteins and which contains at least about 45%

hydrophobic amino acid residues such as alanine, leucine, isoleucine, phenylalanine, proline, tyrosine, tryptophan, or valine. In a preferred embodiment, a signal sequence contains at least about 10 to 35 amino acid residues, preferably about 10 to 20 amino acid residues, and has at least about 35-60%, more preferably 40-50%, and more preferably at least about 45% hydrophobic residues. A signal sequence serves to direct a protein containing such a sequence to a lipid bilayer. Thus, in one embodiment, a TANGO 294 protein contains a signal sequence corresponding to amino acid residues 1 to 33 of SEQ ID NO: 47 (SEQ ID NO: 48). The signal sequence is cleaved during processing of the mature protein.

5

10

15

20

25

The naturally-occurring form of TANGO 294 protein is a secreted protein (i.e., not comprising the predicted signal sequence). However, in variant forms, TANGO 294 proteins can be transmembrane proteins which include an extracellular domain. In this transmembrane variant form, the predicted TANGO 294 protein extracellular domain is located from about amino acid residue 34 to about amino acid residue 254 of SEQ ID NO: 47, the predicted cytoplasmic domain is located from about amino acid residue 280 to amino acid residue 423 of SEQ ID NO: 47 (SEQ ID NO: 52), and the predicted transmembrane domain is located from about amino acid residues 255 to 279 of SEQ ID NO: 47 (SEQ ID NO: 51).

TANGO 294 proteins typically comprise a variety of potential post-translational modification sites (often within an extracellular domain), such as those described herein in Table X, as predicted by computerized sequence analysis of TANGO 294 proteins using amino acid sequence comparison software (comparing the amino acid sequence of TANGO 294 with the information in the PROSITE database {rel. 12.2; Feb, 1995} and the Hidden Markov Models database {Rel. PFAM 3.3}). In certain embodiments, a protein of the invention has at least 1, 2, 4, 6, 10, 15, or 20 or more of the post-translational modification sites listed in Table X.

Table X

| Type of Potential Modification Site | Amino Acid Residues of | Amino Acid |
|---------------------------------------|------------------------|------------|
| or Domain | SEQ ID NO: 47 | Sequence |
| N-glycosylation site | 48 to 51 | NISE |
| | 113 to 116 | NNSL |
| a. | 285 to 288 | NMSR |
| | 413 to 416 | NLSQ |
| Protein kinase C phosphorylation site | 12 to 14 | SHR |
| | 138 to 140 | SRK |
| * | 217 to 219 | TVK |
| Casein kinase II phosphorylation site | 155 to 158 | SYDE |
| | 175 to 178 | TGQE |
| | 198 to 201 | TMPE |
| | 360 to 363 | SNPE |
| Tyrosine kinase phosphorylation site | 174 to 182 | KTGQEKIYY |
| N-myristoylation site | 99 to 104 | GLVGGA |
| | 130 to 135 | GNSRGN |
| | 188 to 193 | GTTMGF |
| · | 277 to 282 | GGFNTN |
| Amidation site | 240 to 243 | FGKK |
| Lipase serine active site | 180 to 189 | IYYVGYSQGT |
| Alpha/beta hydrolase fold domain | 125 to 404 | See Fig. 6 |

Alpha/beta hydrolase fold domains occur in a wide variety of enzymes (Ollis et al., (1992) Protein Eng. 5:197-211). The alpha/beta fold domain 5 is a conserved topological domain in which sequence homology is not necessarily conserved. Conservation of topology in the alpha/beta fold domain preserves arrangement of catalytic residues, even though those residues, and the reactions they catalyze, can vary. In many enzymes, particularly including alpha/beta

hydrolases, this domain encompasses the active site of the enzyme. In one 10

embodiment, the protein of the invention has at least one domain that is at least 55%, preferably at least about 65%, more preferably at least about 75%, yet more preferably at least about 85%, and most preferably at least about 95% identical to the alpha/beta hydrolase fold domain described herein in Table X.

5

10

15

20

25

30

The signal peptide prediction program SIGNALP (Nielsen et al. (1997) *Protein Engineering* 10:1-6) predicted that human TANGO 294 protein includes a 33 amino acid signal peptide (amino acid residues 1 to 33 of SEQ ID NO: 47; SEQ ID NO: 48) preceding the mature TANGO 294 protein (amino acid residues 34 to 423 of SEQ ID NO: 47; SEQ ID NO: 49). Human TANGO 294 protein is a soluble secreted protein. However, in the transmembrane variant form, human TANGO 294 protein includes an extracellular domain (amino acid residues 34 to 254 of SEQ ID NO: 47; SEQ ID NO: 50); a transmembrane domain (amino acid residues 255 to 279 of SEQ ID NO: 47; SEQ ID NO: 51); and a cytoplasmic domain (amino acid residues 280 to 423 of SEQ ID NO: 47; SEQ ID NO: 52).

Figure 6F depicts a hydrophilicity plot of human TANGO 294 protein. Relatively hydrophobic regions are above the dashed horizontal line, and relatively hydrophilic regions are below the dashed horizontal line. The hydrophobic region which corresponds to amino acid residues 1 to 33 of SEQ ID NO: 47 is the signal sequence of human TANGO 294 (SEQ ID NO: 49). The hydrophobic region which corresponds to amino acid residues 255 to 279 of SEQ ID NO: 47 is the predicted transmembrane domain of human TANGO 294 (SEQ ID NO: 51). As described elsewhere herein, relatively hydrophilic regions are generally located at or near the surface of a protein, and are more frequently effective immunogenic epitopes than are relatively hydrophobic regions. For example, the region of human TANGO 294 protein from about amino acid residue 130 to about amino acid residue 150 appears to be located at or near the surface of the protein, while the region from about amino acid residue 90 to about amino acid residue 100 appears not to be located at or near the surface.

The predicted molecular weight of human TANGO 294 protein without modification and prior to cleavage of the signal sequence is about 48.2 kilodaltons. The predicted molecular weight of the mature human TANGO 294

protein without modification and after cleavage of the signal sequence is about 44.2 kilodaltons.

It may be that amino acid residues 1 to 15 of SEQ ID NO: 47 do not occur in TANGO 294 protein. However, it is recognized that amino acid residues 16 to 33 of SEQ ID NO: 47 form a functional signal sequence even in the absence of residues 1 to 15. The amino acid sequence (and hence the properties) of mature TANGO 294 protein are unaffected by presence or absence of amino acid residues 1 to 15 of immature TANGO 294 protein.

5

10

15

20

25

30

Human TANGO 294 protein exhibits considerable sequence similarity (i.e., about 75% amino acid sequence identity) to lingual and gastric lipase proteins of rat (Swissprot Accession no. P04634; Docherty et al. (1985) Nucleic Acids Res. 13:1891-1903), dog (Swissprot Accession no. P80035; Carriere et al. (1991) Eur. J. Biochem. 202:75-83), and human (Swissprot Accession no. P07098; Bernbaeck and Blaeckberg (1987) Biochim. Biophys. Acta 909:237-244), as assessed using the ALIGN v. 2.0 computer software using a pam12.mat scoring matrix and gap penalties of -12/-4. TANGO 294 is distinct from the known human lipase, as indicated in Figures 6D and 6E. Figures 6D and 6E depict an alignment of the amino acid sequences of human TANGO 294 protein (SEQ ID NO: 47) and the known human lipase protein (SEQ ID NO: 75), as assessed using the same software and parameters. In this alignment (pam120.mat scoring matrix, gap penalties -12/-4), the amino acid sequences of the proteins are 49.8% identical. TANGO 294 also is distinct from the known human lysosomal acid lipase, as indicated in Figures 6G and 6H. Figures 6G and 6H depicts an alignment of the amino acid sequences of human TANGO 294 protein (SEQ ID NO: 47) and the known human lysosomal acid lipase protein (SEQ ID NO: 41). In this alignment (pam120.mat scoring matrix, gap penalties -12/-4), the amino acid sequences of the proteins are 56.9% identical.

TANGO 294 is a human lipase distinct from the known human lipase and the known human lysosomal acid lipase. Furthermore, in view of the comparisons of the amino acid sequences of TANGO 294 and the two human lipases and the nature of transcriptional initiation sites, it is recognized that the transcriptional start site can correspond to either of the methionine residues located

at residues 1 and 15 of SEQ ID NO: 47 The present invention thus includes proteins in which the initially transcribed amino acid residue is the methionine residue at position 1 of SEQ ID NO: 47 and proteins in which the initially transcribed amino acid residue is the methionine residue at position 15 of SEQ ID NO: 47 (i.e., proteins in which the amino acid sequence of TANGO 294 does not include residues 1 to 14 of SEQ ID NO: 47). Furthermore, because amino acid residues 1 to 14 of SEQ ID NO: 47 are predicted to be part of a signal sequence, it is recognized that the protein not comprising this portion of the amino acid sequence will nonetheless exhibit a functional signal sequence at its amino terminus.

Biological function of TANGO 294 proteins, nucleic acids, and modulators thereof

10

15

20

25 '

The sequence similarity of TANGO 294 and mammalian lingual, gastric, and lysosomal acid lipase proteins indicates that TANGO 294 is involved in physiological processes identical or analogous to those involving these lipases.

Thus, TANGO 294 is involved in facilitating absorption and metabolism of fat.

TANGO 294 can thus be used, for example, to prevent, detect, and treat disorders relating to fat absorption and metabolism, such as inadequate expression of gastric/pancreatic lipase, cystic fibrosis, exocrine pancreatic insufficiency, obesity, medical treatments which alter fat absorption, and the like.

TANGO 294 protein is known to be expressed in human pulmonary artery smooth muscle tissue. This indicates that TANGO 294 protein is involved in transportation and metabolism of fats and lipids in the human vascular and cardiovascular systems. Thus, TANGO 294 proteins of the invention can be used to prevent, detect, and treat disorders involving these body systems.

INTERCEPT 296

15

20

25

A cDNA clone (designated jthEa030h09) encoding at least a portion of human INTERCEPT 296 protein was isolated from a human esophagus cDNA library. The human INTERCEPT 296 protein is predicted by structural analysis to be a transmembrane protein having three or more transmembrane domains. Expression of DNA encoding INTERCEPT 296 tissue has been detected by northern analysis of human lung tissue. In human lung tissue, two moieties corresponding to INTERCEPT 296 have been identified in Northern blots. It is recognized that these two moieties may represent alternatively polyadenylated INTERCEPT 296 mRNAs or alternatively spliced INTERCEPT 296 mRNAs. It has furthermore been observed that INTERCEPT 296 does not appear to be expressed in any of heart, brain, placenta, skeletal muscle, kidney, and pancreas tissues.

The full length of the cDNA encoding INTERCEPT 296 protein (Figure 7; SEQ ID NO: 53) is 2133 nucleotide residues. The ORF of this cDNA, nucleotide residues 70 to 1098 of SEQ ID NO: 53 (i.e., SEQ ID NO: 54), encodes a 343-amino acid transmembrane protein (Figure 7; SEQ ID NO: 55).

The invention includes purified INTERCEPT 296 protein, which has the amino acid sequence listed in SEQ ID NO: 55. In addition to full length INTERCEPT 296 proteins, the invention includes fragments, derivatives, and variants of these INTERCEPT 296 proteins, as described herein. These proteins, fragments, derivatives, and variants are collectively referred to herein as polypeptides of the invention or proteins of the invention.

The invention also includes nucleic acid molecules which encode a polypeptide of the invention. Such nucleic acids include, for example, a DNA molecule having the nucleotide sequence SEQ ID NO: 53 or some portion thereof, such as the portion which encodes INTERCEPT 296 protein or a domain thereof. These nucleic acids are collectively referred to as nucleic acids of the invention.

INTERCEPT 296 proteins and nucleic acid molecules encoding
them comprise a family of molecules having certain conserved structural and
functional features, such as the five transmembrane domains which occur in the
protein.

5

10

15

20

25

INTERCEPT 296 comprises at least five transmembrane domains, at least three cytoplasmic domains, and at least two extracellular domains. INTERCEPT 296 does not appear to comprise a cleavable signal sequence. Amino acid residues 1 to 70 of SEQ ID NO: 55 likely directs insertion of the protein into the cytoplasmic membrane. There are at least two mechanisms by which this can occur. Sequence analysis of residues 1 to 70 of SEQ ID NO: 55 indicates that this entire region may represent a signal sequence or that residues 1 to 47 represent a signal sequence, with residues 48-70 representing a transmembrane region. Human INTERCEPT 296 protein extracellular domains are located from about amino acid residue 70 to about amino acid residue 182 (SEQ ID NO: 57) and from about amino acid residue 228 to about amino acid residue 249 (SEQ ID NO: 58) of SEQ ID NO: 55. Human INTERCEPT 296 cytoplasmic domains are located from about amino acid residue 43 to amino acid residue 50 (SEQ ID NO: 64), from about amino acid residue 205 to amino acid residue 210 (SEQ ID NO: 65), and from amino acid residue 272 to amino acid residue 343 (SEQ ID NO: 66) of SEQ ID NO: 55. The five transmembrane domains of INTERCEPT 296 are located from about amino acid residues 24 to 42 (SEQ ID NO: 59), 51 to 70 (SEQ ID NO: 60), 183 to 204 (SEQ ID NO: 61), 211 to 227 (SEQ ID NO: 62), and 250 to 271 (SEQ ID NO: 63) of SEQ ID NO: 55.

INTERCEPT 296 proteins typically comprise a variety of potential post-translational modification sites (often within an extracellular domain), such as those described herein in Table XI, as predicted by computerized sequence analysis of INTERCEPT 296 proteins using amino acid sequence comparison software (comparing the amino acid sequence of INTERCEPT 296 with the information in the PROSITE database {rel. 12.2; Feb, 1995} and the Hidden Markov Models database {Rel. PFAM 3.3}). In certain embodiments, a protein of the invention has at least 1, 2, 4, 6, 10, 15, or 20 or more of the post-translational modification sites listed in Table XI.

Table XI

| Type of Potential Modification Site | Amino Acid Residues of | Amino Acid |
|---------------------------------------|------------------------|------------|
| or Domain | SEQ ID NO: 55 | . Sequence |
| N-glycosylation site | 71 to 74 | NFSS |
| | 84 to 87 | NTSY |
| | 109 to 112 | NITL |
| | 121 to 124 | NETI |
| | 284 to 287 | NQSV |
| Protein kinase C phosphorylation site | 86 to 88 | SYK |
| | 131 to 133 | TWR |
| | 162 to 164 | TPR |
| | 304 to 306 | SPR |
| | 313 to 315 | SPK |
| | 326 to 328 | STK |
| Casein kinase II phosphorylation site | 286 to 289 | SVDE |
| 4 | 296 to 299 | SPEE |
| | 309 to 312 | SMAD |
| Tyrosine kinase phosphorylation site | 148 to 156 | KGLPDPVLY |
| γ | | |
| N-myristoylation site | 79 to 84 | GQVSTN |
| | 100 to 105 | GLQVGL |
| | 107 to 112 | GVNITL |
| | 265 to 270 | GLAMAV |

Figure 7D depicts a hydrophilicity plot of INTERCEPT 296 protein.

Relatively hydrophobic regions are above the dashed horizontal line, and relatively hydrophilic regions are below the dashed horizontal line. The hydrophobic regions which corresponds to amino acid residues 24 to 42, 51 to 70, 183 to 204, 211 to 227, and 250 to 271 of SEQ ID NO: 55 are the transmembrane domains of human INTERCEPT 296 (SEQ ID NOs: 59 through 63, respectively). As described

elsewhere herein, relatively hydrophilic regions are generally located at or near the surface of a protein, and are more frequently effective immunogenic epitopes than are relatively hydrophobic regions. For example, the region of human INTERCEPT 296 protein from about amino acid residue 120 to about amino acid residue 140 appears to be located at or near the surface of the protein, while the region from about amino acid residue 95 to about amino acid residue 110 appears not to be located at or near the surface.

5

10

15

20

25

30

The predicted molecular weight of INTERCEPT 296 protein without modification and prior to cleavage of the signal sequence is about 37.8 kilodaltons. The predicted molecular weight of the mature INTERCEPT 296 protein without modification and after cleavage of the signal sequence is about 30.2 kilodaltons.

Figures 7E and 7F depicts an alignment of the amino acid sequences of human INTERCEPT 296 protein (SEQ ID NO: 55) and Caenorhabditis elegans C06E1.3 related protein (SEQ ID NO: 399). In this alignment (pam120.mat scoring matrix, gap penalties -12/-4), the amino acid sequences of the proteins are 26.8% identical. The C. elegans protein has five predicted transmembrane domains.

Biological function of INTERCEPT 296 proteins, nucleic acids, and modulators thereof

The cDNA encoding INTERCEPT 296 protein was obtained from a human esophagus cDNA library, and INTERCEPT 296 is expressed in lung tissue. The INTERCEPT 296-related proteins and nucleic acids of the invention are therefore useful for prevention, detection, and treatment of disorders of the human lung and esophagus. Such disorders include, for example, various cancers, bronchitis, cystic fibrosis, respiratory infections (e.g., influenza, bronchiolitis, pneumonia, and tuberculosis), asthma, emphysema, chronic bronchitis, bronchiectasis, pulmonary edema, pleural effusion, pulmonary embolus, adult and infant respiratory distress syndromes, heartburn, and gastric reflux esophageal disease.

Tables A and B summarize sequence data corresponding to the human proteins herein designated TANGO 202, TANGO 234, TANGO 265, TANGO 273, TANGO 286, TANGO 294, and INTERCEPT 296.

Table A

| Protein | | SEQ ID NOs | | Depicted in | ATCC® |
|-------------|------|------------|---------|-------------|-------------|
| Designation | cDNA | ORF | Protein | Figure # | Accession # |
| TANGO 202 | 1 | 2 | 3 | 1 | 207219 |
| TANGO 234 | 9 | 10 | 11 | 2 | 207184 |
| TANGO 265 | 17 | 18 | 19 | 3 | 207228 |
| TANGO 273 | 25 | 26 | 27 | 4 | 207185 |
| TANGO 286 | 33 | 34 | 35 | 5 | 207220 |
| TANGO 294 | 45 | 46 | 47 | 6 | 207220 |
| INTERCEPT | 53 | 54 | 55 | 7 | 207220 |
| 296 | | | | | |

| B | ĺ |
|----|---|
| 9 | ı |
| -2 | |
| 凸 | 1 |

| Protein Desig. | Signal Sequence | aouen | Mature Protein | ein | Extracellular Domain(s) | я <u> </u> | Transmembrane Domain(s) | ne | Cytoplasmic Domain(s) | o o |
|--------------------------|-----------------|------------|--|------|----------------------------|------------|----------------------------|------|--------------------------|------|
| | | | | | S | SEQ ID NOS | NOs | ľ | | |
| TANGO 202 | 1 to 19 | 4 | 20 to 475 | 2 | 20 to 392 | 9 | 393 to 415 | 7 | 416 to 475 | ∞ |
| (variant) | (1 to 19) | (4) | (20 to 475) | છ | (20 to 475) | 3 | (N/A) | | (N/A) | |
| TANGO 234 | 1 to 40 | 12 | 41 to 1453 | 13 | 41 to 1359 | 14 | 1360 to 1383 | 15 | 1384 to 1453 | 16 |
| TANGO 265 | 1 to 31 | 20 | 32 to 761 | 21 | 32 to 683 | 22 | 684 to 704 | 23 | 705 to 761 | 24 |
| TANGO 273 | 1 to 22 | 28 | 23 to 172 | 29 | 23 to 60 | 30 | 61 to 81 | 31 | 82 to 172 | 32 |
| TANGO 286 | 1 to 23 | 36 | 24 to 455 | 37 | 24 to 455 | 37 | N/A | | N/A | |
| TANGO 294 | 1 to 33 | 48 | 34 to 423 | 49 | 34 to 254 | 20 | 255 to 279 | 51 | 280 to 423 | 52 |
| (variant 1) | (15 to 33) | (40) | (34 to 423) (49) | (49) | (34 to 254) | (20) | (255 to '279) | (51) | (280 to 423) | (52) |
| <variant 2=""></variant> | <1 to 33> | 48 | <1 to 33> <48> <34 to 423> <49> | 499 | <34 to 423> | <49> | <\/\/\> | | <n a=""></n> | |
| {variant 3} | {15 to 33} | {40} | [15 to 33] {40} {34 to 423} {49} {34 to 423} | {49} | {34 to 423} | {49} | { N/A } | | { N/A } | |
| INTERCEPT | N/A | | 1 to 343 | 55 | 1 to 23 | 99 | 24 to 42 | 59 | 43 to 50 | 2 |
| 296 | | | | | 71 to 182 | 57 | 51 to 70 | 09 | 205 to 210 | 65 |
| | | | | | 228 to 249 | 58 | 183 to 204 | 19 | 272 to 343 | 99 |
| | | | | | | | 211 to 227 | 62 | | |
| | | | | | | | 250 to 271 | 63 | | |
| | | | |] | Amino Acid Residues | esidues | | | | |

Various aspects of the invention are described in further detail in the following subsections.

I. Isolated Nucleic Acid Molecules

5

10

15

20

25

30

One aspect of the invention pertains to isolated nucleic acid molecules that encode a polypeptide of the invention or a biologically active portion thereof, as well as nucleic acid molecules sufficient for use as hybridization probes to identify nucleic acid molecules encoding a polypeptide of the invention and fragments of such nucleic acid molecules suitable for use as PCR primers for the amplification or mutation of nucleic acid molecules. As used herein, the term "nucleic acid molecule" is intended to include DNA molecules (e.g., cDNA or genomic DNA) and RNA molecules (e.g., mRNA) and analogs of the DNA or RNA generated using nucleotide analogs. The nucleic acid molecule can be single-stranded or double-stranded, but preferably is double-stranded DNA.

An "isolated" nucleic acid molecule is one which is separated from other nucleic acid molecules which are present in the natural source of the nucleic acid molecule. Preferably, an "isolated" nucleic acid molecule is free of sequences (preferably protein-encoding sequences) which naturally flank the nucleic acid (i.e., sequences located at the 5' and 3' ends of the nucleic acid) in the genomic DNA of the organism from which the nucleic acid is derived. For example, in various embodiments, the isolated nucleic acid molecule can contain less than about 5 kB, 4 kB, 3 kB, 2 kB, 1 kB, 0.5 kB or 0.1 kB of nucleotide sequences which naturally flank the nucleic acid molecule in genomic DNA of the cell from which the nucleic acid is derived. Moreover, an "isolated" nucleic acid molecule, such as a cDNA molecule, can be substantially free of other cellular material, or culture medium when produced by recombinant techniques, or substantially free of chemical precursors or other chemicals when chemically synthesized.

A nucleic acid molecule of the present invention, e.g., a nucleic acid molecule having the nucleotide sequence of all or a portion of any of SEQ ID NOs: 1, 2, 9, 10, 17, 18, 25, 26, 33, 34, 45, 46, 53, 54, 67, 68, 72, and 73, or a complement thereof, or which has a nucleotide sequence comprising one of these sequences, can be isolated using standard molecular biology techniques and the

sequence information provided herein. Using a nucleic acid comprising at least one of the sequences of SEQ ID NOs: 1, 2, 9, 10, 17, 18, 25, 26, 33, 34, 45, 46, 53, 54, 67, 68, 72, and 73 as a hybridization probe, nucleic acid molecules of the invention can be isolated using standard hybridization and cloning techniques (e.g., as described in Sambrook et al., eds., *Molecular Cloning: A Laboratory Manual, 2nd ed., Cold Spring Harbor Laboratory*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989).

5

10

15

20

25

30

A nucleic acid molecule of the invention can be amplified using cDNA, mRNA or genomic DNA as a template and appropriate oligonucleotide primers according to standard PCR amplification techniques. The nucleic acid so amplified can be cloned into an appropriate vector and characterized by DNA sequence analysis. Furthermore, oligonucleotides corresponding to all or a portion of a nucleic acid molecule of the invention can be prepared by standard synthetic techniques, e.g., using an automated DNA synthesizer.

In another preferred embodiment, an isolated nucleic acid molecule of the invention comprises a nucleic acid molecule which is a complement of the nucleotide sequence of any of SEQ ID NOs: 1, 2, 9, 10, 17, 18, 25, 26, 33, 34, 45, 46, 53, 54, 67, 68, 72, and 73, or a portion thereof. A nucleic acid molecule which is complementary to a given nucleotide sequence is one which is sufficiently complementary to the given nucleotide sequence that it can hybridize to the given nucleotide sequence thereby forming a stable duplex.

Moreover, a nucleic acid molecule of the invention can comprise only a portion of a nucleic acid sequence encoding a full length polypeptide of the invention for example, a fragment which can be used as a probe or primer or a fragment encoding a biologically active portion of a polypeptide of the invention. The nucleotide sequence determined from the cloning one gene allows for the generation of probes and primers designed for use in identifying and/or cloning homologs in other cell types, e.g., from other tissues, as well as homologs from other mammals. The probe/primer typically comprises substantially purified oligonucleotide. The oligonucleotide typically comprises a region of nucleotide sequence that hybridizes under stringent conditions to at least about 15, preferably about 25, more preferably about 50, 75, 100, 125, 150, 175, 200, 250, 300, 350, or

400 or more consecutive nucleotides of the sense or anti-sense sequence of any of SEQ ID NOs: 1, 2, 9, 10, 17, 18, 25, 26, 33, 34, 45, 46, 53, 54, 67, 68, 72, and 73, or of a naturally occurring mutant of any of SEQ ID NOs: 1, 2, 9, 10, 17, 18, 25, 26, 33, 34, 45, 46, 53, 54, 67, 68, 72, and 73.

5

10

15

20

25

30

Probes based on the sequence of a nucleic acid molecule of the invention can be used to detect transcripts or genomic sequences encoding the same protein molecule encoded by a selected nucleic acid molecule. The probe comprises a label group attached thereto, e.g., a radioisotope, a fluorescent compound, an enzyme, or an enzyme co-factor. Such probes can be used as part of a diagnostic test kit for identifying cells or tissues which mis-express the protein, such as by measuring levels of a nucleic acid molecule encoding the protein in a sample of cells from a subject, e.g., detecting mRNA levels or determining whether a gene encoding the protein has been mutated or deleted.

A nucleic acid fragment encoding a biologically active portion of a polypeptide of the invention can be prepared by isolating a portion of any of SEQ ID NOs: 2, 10, 18, 26, 34, 46, 54, 68, and 73, expressing the encoded portion of the polypeptide protein (e.g., by recombinant expression *in vitro*), and assessing the activity of the encoded portion of the polypeptide.

The invention further encompasses nucleic acid molecules that differ from the nucleotide sequence of any of SEQ ID NOs: 1, 2, 9, 10, 17, 18, 25, 26, 33, 34, 45, 46, 53, 54, 67, 68, 72, and 73 due to degeneracy of the genetic code and thus encode the same protein as that encoded by the nucleotide sequence of any of SEQ ID NOs: 2, 10, 18, 26, 34, 46, 54, 68, and 73.

In addition to the nucleotide sequences of SEQ ID NOs: 2, 10, 18, 26, 34, 46, 54, 68, and 73, it will be appreciated by those skilled in the art that DNA sequence polymorphisms that lead to changes in the amino acid sequence can exist within a population (e.g., the human population). Such genetic polymorphisms can exist among individuals within a population due to natural allelic variation. An allele is one of a group of genes which occur alternatively at a given genetic locus.

As used herein, the phrase "allelic variant" refers to a nucleotide sequence which occurs at a given locus or to a polypeptide encoded by the nucleotide sequence. For example, chromosomal mapping has been used to locate

the gene encoding human TANGO 234 at chromosomal location h12p13 (with synteny to mo6), between chromosomal markers WI-6980 and GATA8A09.43. Thus, human TANGO 234 allelic variants can include TANGO 234 nucleotide sequence polymorphisms (e.g., nucleotide sequences that vary from SEQ ID NO: 9) that map to this chromosomal region. Similarly, chromosomal mapping has been used to locate the gene encoding human TANGO 265 protein on chromosome 1, between markers D1S305 and D1S2635. Allelic variants of TANGO 265 occur at this chromosomal location. Further by way of example, the gene encoding human TANGO 273 protein has been located by chromosomal mapping on chromosome 7, between markers D7S2467 and D7S2552. Allelic variants of TANGO 273 occur at this chromosomal location.

5

10

15

20

25

30

As used herein, the terms "gene" and "recombinant gene" refer to nucleic acid molecules comprising an open reading frame encoding a polypeptide of the invention. Such natural allelic variations can typically result in 1-5% variance in the nucleotide sequence of a given gene. Alternative alleles can be identified by sequencing the gene of interest in a number of different individuals. This can be readily carried out by using hybridization probes to identify the same genetic locus in a variety of individuals. Any and all such nucleotide variations and resulting amino acid polymorphisms or variations that are the result of natural allelic variation and that do not alter the functional activity are intended to be within the scope of the invention.

Moreover, nucleic acid molecules encoding proteins of the invention from other species (homologs), which have a nucleotide sequence which differs from that of the specific proteins described herein are intended to be within the scope of the invention. Nucleic acid molecules corresponding to natural allelic variants and homologs of a cDNA of the invention can be isolated based on their homology with nucleic acid molecules described herein, using the specific cDNAs described herein, or a portion thereof, as a hybridization probe according to standard hybridization techniques under stringent hybridization conditions. For example, a cDNA encoding a soluble form of a membrane-bound protein of the invention isolated based on its hybridization to a nucleic acid molecule encoding all or part of the membrane-bound form. Likewise, a cDNA encoding a membrane-

bound form can be isolated based on its hybridization to a nucleic acid molecule encoding all or part of the soluble form.

5

10

15

20

25

30

Accordingly, in another embodiment, an isolated nucleic acid molecule of the invention is at least 15 (25, 40, 60, 80, 100, 150, 200, 250, 300, 350, 400, 450, 550, 650, 700, 800, 900, 1000, 1200, 1400, 1600, 1800, 2000, 2200, 2400, 2600, 2800, 3000, 3500, 4000, 4500, or 4928) nucleotides in length and hybridizes under stringent conditions to the nucleic acid molecule comprising the nucleotide sequence, preferably the coding sequence, of any of SEQ ID NOs: 1, 2, 9, 10, 17, 18, 25, 26, 33, 34, 45, 46, 53, 54, 67, 68, 72, and 73, or a complement thereof. As used herein, the term "hybridizes under stringent conditions" is intended to describe conditions for hybridization and washing under which nucleotide sequences at least 60% (65%, 70%, preferably 75%) identical to each other typically remain hybridized to each other. Such stringent conditions are known to those skilled in the art and can be found in Current Protocols in Molecular Biology, John Wiley & Sons, N.Y. (1989), 6.3.1-6.3.6. A preferred, nonlimiting example of stringent hybridization conditions are hybridization in 6 × sodium chloride/sodium citrate (SSC) at about 45°C, followed by one or more washes in 0.2 × SSC, 0.1% SDS at 50-65°C. Preferably, an isolated nucleic acid molecule of the invention that hybridizes under stringent conditions with the sequence of any of SEQ ID NOs: 1, 2, 9, 10, 17, 18, 25, 26, 33, 34, 45, 46, 53, 54, 67, 68, 72, and 73, or a complement thereof, corresponds to a naturally-occurring nucleic acid molecule. As used herein, a "naturally-occurring" nucleic acid molecule refers to an RNA or DNA molecule having a nucleotide sequence that occurs in nature (e.g., encodes a natural protein).

In addition to naturally-occurring allelic variants of a nucleic acid molecule of the invention sequence that can exist in the population, the skilled artisan will further appreciate that changes can be introduced by mutation thereby leading to changes in the amino acid sequence of the encoded protein, without altering the biological activity of the protein. For example, one can make nucleotide substitutions leading to amino acid substitutions at "non-essential" amino acid residues. A "non-essential" amino acid residue is a residue that can be altered from the wild-type sequence without altering the biological activity, whereas

an "essential" amino acid residue is required for biological activity. For example, amino acid residues that are not conserved or only semi-conserved among homologs of various species may be non-essential for activity and thus would be likely targets for alteration. Alternatively, amino acid residues that are conserved among the homologs of various species (e.g., murine and human) may be essential for activity and thus would not be likely targets for alteration.

5

10

15

20

25

30

Accordingly, another aspect of the invention pertains to nucleic acid molecules encoding a polypeptide of the invention that contain changes in amino acid residues that are not essential for activity. Such polypeptides differ in amino acid sequence from the sequence of any of SEQ ID NOs: 3-8, 11-16, 19-24, 27-32, 35-44, 47-52, 55-66, 69, and 74, yet retain biological activity. In one embodiment, the isolated nucleic acid molecule includes a nucleotide sequence encoding a protein that includes an amino acid sequence that is at least about 40% identical, 50%, 60%, 70%, 80%, 90%, 95%, or 98% identical to the amino acid sequence of any of SEQ ID NOs: 3-8, 11-16, 19-24, 27-32, 35-44, 47-52, 55-66, 69, and 74.

An isolated nucleic acid molecule encoding a variant protein can be created by introducing one or more nucleotide substitutions, additions or deletions into the nucleotide sequence of any of SEQ ID NOs: 1, 2, 9, 10, 17, 18, 25, 26, 33, 34, 45, 46, 53, 54, 67, 68, 72, and 73, such that one or more amino acid residue substitutions, additions or deletions are introduced into the encoded protein. Mutations can be introduced by standard techniques, such as site-directed mutagenesis and PCR-mediated mutagenesis. Preferably, conservative amino acid substitutions are made at one or more predicted non-essential amino acid residues. A "conservative amino acid substitution" is one in which the amino acid residue is replaced with an amino acid residue having a similar side chain. Families of amino acid residues having similar side chains have been defined in the art. These families include amino acids with basic side chains (e.g., lysine, arginine, histidine), acidic side chains (e.g., aspartic acid, glutamic acid), uncharged polar side chains (e.g., glycine, asparagine, glutamine, serine, threonine, tyrosine, cysteine), nonpolar side chains (e.g., alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine, tryptophan), beta-branched side chains (e.g., threonine, valine, isoleucine) and aromatic side chains (e.g., tyrosine, phenylalanine, tryptophan,

histidine). Alternatively, mutations can be introduced randomly along all or part of the coding sequence, such as by saturation mutagenesis, and the resultant mutants can be screened for biological activity to identify mutants that retain activity. Following mutagenesis, the encoded protein can be expressed recombinantly and the activity of the protein can be determined.

5

10

15

20

25

30

In a preferred embodiment, a mutant polypeptide that is a variant of a polypeptide of the invention can be assayed for: (1) the ability to form protein:protein interactions with one or more polypeptides of the invention (e.g., in a signaling pathway); (2) the ability to bind a ligand of a polypeptide of the invention (e.g., another protein identified herein); (3) the ability to bind to an intracellular target protein of a polypeptide of the invention (e.g., a modulator or substrate of the polypeptide); or (4) the ability to modulate a physiological activity of the protein, such as one of those disclosed herein (e.g., ability to modulate cell proliferation, cell migration, chemotaxis, or cellular differentiation).

The present invention encompasses antisense nucleic acid molecules, i.e., molecules which are complementary to a sense nucleic acid encoding a polypeptide of the invention, e.g., complementary to the coding strand of a double-stranded cDNA molecule or complementary to an mRNA sequence. Accordingly, an antisense nucleic acid can hydrogen bond to a sense nucleic acid. The antisense nucleic acid can be complementary to an entire coding strand, or to only a portion thereof, e.g., all or part of the protein coding region (or open reading frame). An antisense nucleic acid molecule can be antisense to all or part of a non-coding region of the coding strand of a nucleotide sequence encoding a polypeptide of the invention. The non-coding regions ("5' and 3' untranslated regions") are the 5' and 3' sequences which flank the coding region and are not translated into amino acids.

An antisense oligonucleotide can be, for example, about 5, 10, 15, 20, 25, 30, 35, 40, 45, or 50 or more nucleotides in length. An antisense nucleic acid of the invention can be constructed using chemical synthesis and enzymatic ligation reactions using procedures known in the art. For example, an antisense nucleic acid (e.g., an antisense oligonucleotide) can be chemically synthesized using naturally occurring nucleotides or variously modified nucleotides designed to increase the biological stability of the molecules or to increase the physical stability

of the duplex formed between the antisense and sense nucleic acids, e.g., phosphorothioate derivatives and acridine substituted nucleotides can be used. Examples of modified nucleotides which can be used to generate the antisense nucleic acid include 5-fluorouracil, 5-bromouracil, 5-chlorouracil, 5-iodouracil, hypoxanthine, xanthine, 4-acetylcytosine, 5-(carboxyhydroxylmethyl) uracil, 5carboxymethylaminomethyl-2-thiouridine, 5-carboxymethylaminomethyluracil, dihydrouracil, beta-D-galactosylqueosine, inosine, N₆-isopentenyladenine, 1methylguanine, 1-methylinosine, 2,2-dimethylguanine, 2-methyladenine, 2methylguanine, 3-methylcytosine, 5-methylcytosine, N6-adenine, 7-methylguanine, 5-methylaminomethyluracil, 5-methoxyaminomethyl-2-thiouracil, beta-Dmannosylqueosine, 5'-methoxycarboxymethyluracil, 5-methoxyuracil, 2methylthio-N6-isopentenyladenine, uracil-5-oxyacetic acid (v), wybutoxosine, pseudouracil, queosine, 2-thiocytosine, 5-methyl-2-thiouracil, 2-thiouracil, 4thiouracil, 5-methyluracil, uracil-5-oxyacetic acid methylester, uracil-5-oxyacetic acid (v), 5-methyl-2-thiouracil, 3-(3-amino-3-N-2-carboxypropyl) uracil, (acp3)w, and 2,6-diaminopurine. Alternatively, the antisense nucleic acid can be produced biologically using an expression vector into which a nucleic acid has been subcloned in an antisense orientation (i.e., RNA transcribed from the inserted nucleic acid will be of an antisense orientation to a target nucleic acid of interest, described further in the following subsection).

10

15

20

25

30

The antisense nucleic acid molecules of the invention are typically administered to a subject or generated *in situ* such that they hybridize with or bind to cellular mRNA and/or genomic DNA encoding a selected polypeptide of the invention to thereby inhibit expression, e.g., by inhibiting transcription and/or translation. The hybridization can be by conventional nucleotide complementarity to form a stable duplex, or, for example, in the case of an antisense nucleic acid molecule which binds to DNA duplexes, through specific interactions in the major groove of the double helix. An example of a route of administration of antisense nucleic acid molecules of the invention includes direct injection at a tissue site. Alternatively, antisense nucleic acid molecules can be modified to target selected cells and then administered systemically. For example, for systemic administration, antisense molecules can be modified such that they specifically bind to receptors or

antigens expressed on a selected cell surface, e.g., by linking the antisense nucleic acid molecules to peptides or antibodies which bind to cell surface receptors or antigens. The antisense nucleic acid molecules can also be delivered to cells using the vectors described herein. To achieve sufficient intracellular concentrations of the antisense molecules, vector constructs in which the antisense nucleic acid molecule is placed under the control of a strong pol II or pol III promoter are preferred.

5

10

15

20

25

30

An antisense nucleic acid molecule of the invention can be an alphaanomeric nucleic acid molecule. An α-anomeric nucleic acid molecule forms specific double-stranded hybrids with complementary RNA in which, contrary to the usual beta-units, the strands run parallel to each other (Gaultier et al. (1987) Nucleic Acids Res. 15:6625-6641). The antisense nucleic acid molecule can also comprise a 2'-o-methylribonucleotide (Inoue et al. (1987) Nucleic Acids Res. 15:6131-6148) or a chimeric RNA-DNA analogue (Inoue et al. (1987) FEBS Lett. 215:327-330).

The invention also encompasses ribozymes. Ribozymes are catalytic RNA molecules with ribonuclease activity which are capable of cleaving a single-stranded nucleic acid, such as an mRNA, to which they have a complementary region. Thus, ribozymes (e.g., hammerhead ribozymes as described in Haselhoff and Gerlach (1988) *Nature* 334:585-591) can be used to catalytically cleave mRNA transcripts to thereby inhibit translation of the protein encoded by the mRNA. A ribozyme having specificity for a nucleic acid molecule encoding a polypeptide of the invention can be designed based upon the nucleotide sequence of a cDNA disclosed herein. For example, a derivative of a *Tetrahymena* L-19 IVS RNA can be constructed in which the nucleotide sequence of the active site is complementary to the nucleotide sequence to be cleaved in a Cech et al. U.S. Patent No. 4,987,071; and Cech et al. U.S. Patent No. 5,116,742. Alternatively, an mRNA encoding a polypeptide of the invention can be used to select a catalytic RNA having a specific ribonuclease activity from a pool of RNA molecules. *See, e.g.*, Bartel and Szostak (1993) *Science* 261:1411-1418.

The invention also encompasses nucleic acid molecules which form triple helical structures. For example, expression of a polypeptide of the invention

can be inhibited by targeting nucleotide sequences complementary to the regulatory region of the gene encoding the polypeptide (e.g., the promoter and/or enhancer) to form triple helical structures that prevent transcription of the gene in target cells.

See generally Helene (1991) Anticancer Drug Des. 6(6):569-84; Helene (1992)

Ann. N.Y. Acad. Sci. 660:27-36; and Maher (1992) Bioassays 14(12):807-15.

5

10

15

20

25

30

In various embodiments, the nucleic acid molecules of the invention can be modified at the base moiety, sugar moiety or phosphate backbone to improve, e.g., the stability, hybridization, or solubility of the molecule. For example, the deoxyribose phosphate backbone of the nucleic acids can be modified to generate peptide nucleic acids (see Hyrup et al. (1996) Bioorganic & Medicinal Chemistry 4(1): 5-23). As used herein, the terms "peptide nucleic acids" or "PNAs" refer to nucleic acid mimics, e.g., DNA mimics, in which the deoxyribose phosphate backbone is replaced by a pseudopeptide backbone and only the four natural nucleobases are retained. The neutral backbone of PNAs has been shown to allow for specific hybridization to DNA and RNA under conditions of low ionic strength. The synthesis of PNA oligomers can be performed using standard solid phase peptide synthesis protocols as described in Hyrup et al. (1996), supra; Perry-O'Keefe et al. (1996) Proc. Natl. Acad. Sci. USA 93: 14670-675.

PNAs can be used in therapeutic and diagnostic applications. For example, PNAs can be used as antisense or anti-gene agents for sequence-specific modulation of gene expression by, e.g., inducing transcription or translation arrest or inhibiting replication. PNAs can also be used, e.g., in the analysis of single base pair mutations in a gene by, e.g., PNA directed PCR clamping; as artificial restriction enzymes when used in combination with other enzymes, e.g., S1 nucleases (Hyrup (1996), *supra*; or as probes or primers for DNA sequence and hybridization (Hyrup (1996), *supra*; Perry-O'Keefe et al. (1996) *Proc. Natl. Acad. Sci. USA* 93: 14670-675).

In another embodiment, PNAs can be modified, e.g., to enhance their stability or cellular uptake, by attaching lipophilic or other helper groups to PNA, by the formation of PNA-DNA chimeras, or by the use of liposomes or other techniques of drug delivery known in the art. For example, PNA-DNA chimeras can be generated which can combine the advantageous properties of PNA and

DNA. Such chimeras allow DNA recognition enzymes, e.g., RNase H and DNA polymerases, to interact with the DNA portion while the PNA portion would provide high binding affinity and specificity. PNA-DNA chimeras can be linked using linkers of appropriate lengths selected in terms of base stacking, number of bonds between the nucleobases, and orientation (Hyrup (1996), supra). The synthesis of PNA-DNA chimeras can be performed as described in Hyrup (1996). supra, and Finn et al. (1996) Nucleic Acids Res. 24(17):3357-63. For example, a DNA chain can be synthesized on a solid support using standard phosphoramidite coupling chemistry and modified nucleoside analogs. Compounds such as 5'-(4methoxytrityl)amino-5'-deoxy-thymidine phosphoramidite can be used as a link between the PNA and the 5' end of DNA (Mag et al. (1989) Nucleic Acids Res. 17:5973-88). PNA monomers are then coupled in a step-wise manner to produce a chimeric molecule with a 5' PNA segment and a 3' DNA segment (Finn et al. (1996) Nucleic Acids Res. 24(17):3357-63). Alternatively, chimeric molecules can be synthesized with a 5' DNA segment and a 3' PNA segment (Peterser et al. (1975) Bioorganic Med. Chem. Lett. 5:1119-11124).

In other embodiments, the oligonucleotide can include other appended groups such as peptides (e.g., for targeting host cell receptors in vivo), or agents facilitating transport across the cell membrane (see, e.g., Letsinger et al. (1989) Proc. Natl. Acad. Sci. USA 86:6553-6556; Lemaitre et al. (1987) Proc. Natl. Acad. Sci. USA 84:648-652; PCT Publication No. WO 88/09810) or the blood-brain barrier (see, e.g., PCT Publication No. WO 89/10134). In addition, oligonucleotides can be modified with hybridization-triggered cleavage agents (see, e.g., Krol et al. (1988) Bio/Techniques 6:958-976) or intercalating agents (see, e.g., Zon (1988) Pharm. Res. 5:539-549). To this end, the oligonucleotide can be conjugated to another molecule, e.g., a peptide, hybridization triggered cross-linking agent, transport agent, hybridization-triggered cleavage agent, etc.

II. Isolated Proteins and Antibodies

5

10

15

20

25

30

One aspect of the invention pertains to isolated proteins, and biologically active portions thereof, as well as polypeptide fragments suitable for use as immunogens to raise antibodies directed against a polypeptide of the

invention. In one embodiment, the native polypeptide can be isolated from cells or tissue sources by an appropriate purification scheme using standard protein purification techniques. In another embodiment, polypeptides of the invention are produced by recombinant DNA techniques. Alternative to recombinant expression, a polypeptide of the invention can be synthesized chemically using standard peptide synthesis techniques.

5

10

15

20

25

30

An "isolated" or "purified" protein or biologically active portion thereof is substantially free of cellular material or other contaminating proteins from the cell or tissue source from which the protein is derived, or substantially free of chemical precursors or other chemicals when chemically synthesized. The language "substantially free of cellular material" includes preparations of protein in which the protein is separated from cellular components of the cells from which it is isolated or recombinantly produced. Thus, protein that is substantially free of cellular material includes preparations of protein having less than about 30%, 20%, 10%, or 5% (by dry weight) of heterologous protein (also referred to herein as a "contaminating protein"). When the protein or biologically active portion thereof is recombinantly produced, it is also preferably substantially free of culture medium, i.e., culture medium represents less than about 20%, 10%, or 5% of the volume of the protein preparation. When the protein is produced by chemical synthesis, it is preferably substantially free of chemical precursors or other chemicals, i.e., it is separated from chemical precursors or other chemicals which are involved in the synthesis of the protein. Accordingly such preparations of the protein have less than about 30%, 20%, 10%, 5% (by dry weight) of chemical precursors or compounds other than the polypeptide of interest.

Biologically active portions of a polypeptide of the invention include polypeptides comprising amino acid sequences sufficiently identical to or derived from the amino acid sequence of the protein (e.g., the amino acid sequence shown in any of SEQ ID NOs: 3-8, 11-16, 19-24, 27-32, 35-44, 47-52, 55-66, 69, and 74), which include fewer amino acids than the full length protein, and exhibit at least one activity of the corresponding full-length protein. Typically, biologically active portions comprise a domain or motif with at least one activity of the corresponding protein. A biologically active portion of a protein of the invention can be a

polypeptide which is, for example, 10, 25, 50, 100 or more amino acids in length. Moreover, other biologically active portions, in which other regions of the protein are deleted, can be prepared by recombinant techniques and evaluated for one or more of the functional activities of the native form of a polypeptide of the invention.

5

10

15

20

25

30

Preferred polypeptides have the amino acid sequence of any of SEQ ID NOs: 3-8, 11-16, 19-24, 27-32, 35-44, 47-52, 55-66, 69, and 74. Other useful proteins are substantially identical (e.g., at least about 40%, preferably 50%, 60%, 70%, 80%, 90%, 95%, or 99%) to any of SEQ ID NOs: 3-8, 11-16, 19-24, 27-32, 35-44, 47-52, 55-66, 69, and 74 and retain the functional activity of the protein of the corresponding naturally-occurring protein yet differ in amino acid sequence due to natural allelic variation or mutagenesis.

To determine the percent identity of two amino acid sequences or of two nucleic acids, the sequences are aligned for optimal comparison purposes (e.g., gaps can be introduced in the sequence of a first amino acid or nucleic acid sequence for optimal alignment with a second amino or nucleic acid sequence). The amino acid residues or nucleotides at corresponding amino acid positions or nucleotide positions are then compared. When a position in the first sequence is occupied by the same amino acid residue or nucleotide as the corresponding position in the second sequence, then the molecules are identical at that position. The percent identity between the two sequences is a function of the number of identical positions shared by the sequences (i.e., % identity = # of identical positions/total # of positions (e.g., overlapping positions) × 100). In one embodiment the two sequences are the same length.

The determination of percent identity between two sequences can be accomplished using a mathematical algorithm. A preferred, non-limiting example of a mathematical algorithm utilized for the comparison of two sequences is the algorithm of Karlin and Altschul (1990) *Proc. Natl. Acad. Sci. USA* 87:2264-2268, modified as in Karlin and Altschul (1993) *Proc. Natl. Acad. Sci. USA* 90:5873-5877. Such an algorithm is incorporated into the NBLAST and XBLAST programs of Altschul, et al. (1990) *J. Mol. Biol.* 215:403-410. BLAST nucleotide searches can be performed with the NBLAST program, score = 100, wordlength = 12 to

obtain nucleotide sequences homologous to a nucleic acid molecules of the invention. BLAST protein searches can be performed with the XBLAST program, score = 50, wordlength = 3 to obtain amino acid sequences homologous to a protein molecules of the invention. To obtain gapped alignments for comparison purposes, Gapped BLAST can be utilized as described in Altschul et al. (1997) Nucleic Acids Res. 25:3389-3402. Alternatively, PSI-Blast can be used to perform an iterated search which detects distant relationships between molecules. Id. When utilizing BLAST, Gapped BLAST, and PSI-Blast programs, the default parameters of the respective programs (e.g., XBLAST and NBLAST) can be used. See http://www.ncbi.nlm.nih.gov. Another preferred, non-limiting example of a mathematical algorithm utilized for the comparison of sequences is the algorithm of Myers and Miller, (1988) CABIOS 4:11-17. Such an algorithm is incorporated into the ALIGN program (version 2.0) which is part of the GCG sequence alignment software package. When utilizing the ALIGN program for comparing amino acid sequences, a PAM120 weight residue table, a gap length penalty of 12, and a gap penalty of 4 can be used.

5

10

15

20

25

30

The percent identity between two sequences can be determined using techniques similar to those described above, with or without allowing gaps. In calculating percent identity, only exact matches are counted.

The invention also provides chimeric or fusion proteins. As used herein, a "chimeric protein" or "fusion protein" comprises all or part (preferably biologically active) of a polypeptide of the invention operably linked to a heterologous polypeptide (i.e., a polypeptide other than the same polypeptide of the invention). Within the fusion protein, the term "operably linked" is intended to indicate that the polypeptide of the invention and the heterologous polypeptide are fused in-frame to each other. The heterologous polypeptide can be fused to the amino-terminus or the carboxyl-terminus of the polypeptide of the invention.

One useful fusion protein is a GST fusion protein in which the polypeptide of the invention is fused to the carboxyl terminus of GST sequences. Such fusion proteins can facilitate the purification of a recombinant polypeptide of the invention.

In another embodiment, the fusion protein contains a heterologous signal sequence at its amino terminus. For example, the native signal sequence of a polypeptide of the invention can be removed and replaced with a signal sequence from another protein. For example, the gp67 secretory sequence of the baculovirus envelope protein can be used as a heterologous signal sequence (Current Protocols in Molecular Biology, Ausubel et al., eds., John Wiley & Sons, 1992). Other examples of eukaryotic heterologous signal sequences include the secretory sequences of melittin and human placental alkaline phosphatase (Stratagene; La Jolla, California). In yet another example, useful prokaryotic heterologous signal sequences include the phoA secretory signal (Sambrook et al., supra) and the protein A secretory signal (Pharmacia Biotech; Piscataway, New Jersey).

5

10

15

20

25

30

In yet another embodiment, the fusion protein is an immunoglobulin fusion protein in which all or part of a polypeptide of the invention is fused to sequences derived from a member of the immunoglobulin protein family. The immunoglobulin fusion proteins of the invention can be incorporated into pharmaceutical compositions and administered to a subject to inhibit an interaction between a ligand (soluble or membrane-bound) and a protein on the surface of a cell (receptor), to thereby suppress signal transduction *in vivo*. The immunoglobulin fusion protein can be used to affect the bioavailability of a cognate ligand of a polypeptide of the invention. Inhibition of ligand/receptor interaction can be useful therapeutically, both for treating proliferative and differentiative disorders and for modulating (e.g., promoting or inhibiting) cell survival. Moreover, the immunoglobulin fusion proteins of the invention can be used as immunogens to produce antibodies directed against a polypeptide of the invention in a subject, to purify ligands and in screening assays to identify molecules which inhibit the interaction of receptors with ligands.

Chimeric and fusion proteins of the invention can be produced by standard recombinant DNA techniques. In another embodiment, the fusion gene can be synthesized by conventional techniques including automated DNA synthesizers. Alternatively, PCR amplification of gene fragments can be carried out using anchor primers which give rise to complementary overhangs between two consecutive gene fragments which can subsequently be annealed and re-amplified

to generate a chimeric gene sequence (see, e.g., Ausubel et al., supra). Moreover, many expression vectors are commercially available that already encode a fusion moiety (e.g., a GST polypeptide). A nucleic acid encoding a polypeptide of the invention can be cloned into such an expression vector such that the fusion moiety is linked in-frame to the polypeptide of the invention.

5

10

15

20

25

30

A signal sequence of a polypeptide of the invention (e.g., the signal sequence in one of SEQ ID NOs: 3, 4, 11, 12, 19, 20, 27, 28, 35, 36, 47, 48, 69, and 74) can be used to facilitate secretion and isolation of the secreted protein or other proteins of interest. Signal sequences are typically characterized by a core of hydrophobic amino acids which are generally cleaved from the mature protein during secretion in one or more cleavage events. Such signal peptides contain processing sites that allow cleavage of the signal sequence from the mature proteins as they pass through the secretory pathway. Thus, the invention pertains to the described polypeptides having a signal sequence, as well as to the signal sequence itself and to the polypeptide in the absence of the signal sequence (i.e., the cleavage products). In one embodiment, a nucleic acid sequence encoding a signal sequence of the invention can be operably linked in an expression vector to a protein of interest, such as a protein which is ordinarily not secreted or is otherwise difficult to isolate. The signal sequence directs secretion of the protein, such as from a eukaryotic host into which the expression vector is transformed, and the signal sequence is subsequently or concurrently cleaved. The protein can then be readily purified from the extracellular medium by art recognized methods. Alternatively, the signal sequence can be linked to the protein of interest using a sequence which facilitates purification, such as with a GST domain.

In another embodiment, the signal sequences of the present invention can be used to identify regulatory sequences, e.g., promoters, enhancers, repressors. Since signal sequences are the most amino-terminal sequences of a peptide, it is expected that the nucleic acids which flank the signal sequence on its amino-terminal side will be regulatory sequences which affect transcription. Thus, a nucleotide sequence which encodes all or a portion of a signal sequence can be used as a probe to identify and isolate signal sequences and their flanking regions, and these flanking regions can be studied to identify regulatory elements therein.

The present invention also pertains to variants of the polypeptides of the invention. Such variants have an altered amino acid sequence which can function as either agonists (mimetics) or as antagonists. Variants can be generated by mutagenesis, e.g., discrete point mutation or truncation. An agonist can retain substantially the same, or a subset, of the biological activities of the naturally occurring form of the protein. An antagonist of a protein can inhibit one or more of the activities of the naturally occurring form of the protein by, for example, competitively binding to a downstream or upstream member of a cellular signaling cascade which includes the protein of interest. Thus, specific biological effects can be elicited by treatment with a variant of limited function. Treatment of a subject with a variant having a subset of the biological activities of the naturally occurring form of the protein can have fewer side effects in a subject relative to treatment with the naturally occurring form of the protein.

5

10

15

20

25

30

Variants of a protein of the invention which function as either agonists (mimetics) or as antagonists can be identified by screening combinatorial libraries of mutants, e.g., truncation mutants, of the protein of the invention for agonist or antagonist activity. In one embodiment, a variegated library of variants is generated by combinatorial mutagenesis at the nucleic acid level and is encoded by a variegated gene library. A variegated library of variants can be produced by, for example, enzymatically ligating a mixture of synthetic oligonucleotides into gene sequences such that a degenerate set of potential protein sequences is expressible as individual polypeptides, or alternatively, as a set of larger fusion proteins (e.g., for phage display). There are a variety of methods which can be used to produce libraries of potential variants of the polypeptides of the invention from a degenerate oligonucleotide sequence. Methods for synthesizing degenerate oligonucleotides are known in the art (see, e.g., Narang (1983) Tetrahedron 39:3; Itakura et al. (1984) Annu. Rev. Biochem. 53:323; Itakura et al. (1984) Science 198:1056; Ike et al. (1983) Nucleic Acid Res. 11:477).

In addition, libraries of fragments of the coding sequence of a polypeptide of the invention can be used to generate a variegated population of polypeptides for screening and subsequent selection of variants. For example, a library of coding sequence fragments can be generated by treating a double stranded

PCR fragment of the coding sequence of interest with a nuclease under conditions wherein nicking occurs only about once per molecule, denaturing the double stranded DNA, re-naturing the DNA to form double stranded DNA which can include sense/antisense pairs from different nicked products, removing single stranded portions from reformed duplexes by treatment with S1 nuclease, and ligating the resulting fragment library into an expression vector. By this method, an expression library can be derived which encodes amino terminal and internal fragments of various sizes of the protein of interest.

5

10

15

20

25

30

Several techniques are known in the art for screening gene products of combinatorial libraries made by point mutations or truncation, and for screening cDNA libraries for gene products having a selected property. The most widely used techniques, which are amenable to high through-put analysis, for screening large gene libraries typically include cloning the gene library into replicable expression vectors, transforming appropriate cells with the resulting library of vectors, and expressing the combinatorial genes under conditions in which detection of a desired activity facilitates isolation of the vector encoding the gene whose product was detected. Recursive ensemble mutagenesis (REM), a technique which enhances the frequency of functional mutants in the libraries, can be used in combination with the screening assays to identify variants of a protein of the invention (Arkin and Yourvan (1992) *Proc. Natl. Acad. Sci. USA* 89:7811-7815; Delgrave et al. (1993) *Protein Engineering* 6(3):327-331).

An isolated polypeptide of the invention, or a fragment thereof, can be used as an immunogen to generate antibodies using standard techniques for polyclonal and monoclonal antibody preparation. The full-length polypeptide or protein can be used or, alternatively, the invention provides antigenic peptide fragments for use as immunogens. The antigenic peptide of a protein of the invention comprises at least 8 (preferably 10, 15, 20, or 30 or more) amino acid residues of the amino acid sequence of any of SEQ ID NOs: 3-8, 11-16, 19-24, 27-32, 35-44, 47-52, 55-66, 69, and 74, and encompasses an epitope of the protein such that an antibody raised against the peptide forms a specific immune complex with the protein.

preparations are ones that contain only antibodies directed against one or more polypeptides of the invention. Particularly preferred immunogen compositions are those that contain no other human proteins such as, for example, immunogen compositions made using a non-human host cell for recombinant expression of a polypeptide of the invention. In such a manner, the only human epitope or epitopes recognized by the resulting antibody compositions raised against this immunogen will be present as part of a polypeptide or polypeptides of the invention.

5

10

15

20

25

30

The antibody titer in the immunized subject can be monitored over time by standard techniques, such as with an enzyme linked immunosorbent assay (ELISA) using immobilized polypeptide. If desired, the antibody molecules can be harvested or isolated from the subject (e.g., from the blood or serum of the subject) and further purified by well-known techniques, such as protein A chromatography to obtain the IgG fraction. Alternatively, antibodies which bind specifically with a protein or polypeptide of the invention can be selected or purified (e.g., partially purified) using chromatographic methods, such as affinity chromatography. For example, a recombinantly expressed and purified (or partially purified) protein of the invention can be produced as described herein, and covalently or non-covalently coupled with a solid support such as, for example, a chromatography column. The column thus exhibits specific affinity for antibody substances which bind specifically with the protein of the invention, and these antibody substances can be purified from a sample containing antibody substances directed against a large number of different epitopes, thereby generating a substantially purified antibody substance composition, i.e., one that is substantially free of antibody substances which do not bind specifically with the protein. A substantially purified antibody composition, in this context, means an antibody sample that contains at most only 30% (by dry weight) of contaminating antibodies directed against epitopes other than those on the desired protein or polypeptide of the invention, preferably at most 20%, more preferably at most 10%, most preferably at most 5% (by dry weight of the sample is contaminating antibodies). A purified antibody composition means that at least 99% of the antibodies in the composition are directed against the desired protein or polypeptide of the invention.

Preferred epitopes encompassed by the antigenic peptide are regions that are located on the surface of the protein, e.g., hydrophilic regions. Figures 1L, 1M, 2J, 3U, 4I, 4J, 5E, 6F, and 7D are hydrophobicity plots of the proteins of the invention. These plots or similar analyses can be used to identify hydrophilic regions.

5

10

15

20

25

30

An immunogen typically is used to prepare antibodies by immunizing a suitable (i.e., immunocompetent) subject such as a rabbit, goat, mouse, or other mammal or vertebrate. An appropriate immunogenic preparation can contain, for example, recombinantly-expressed or chemically-synthesized polypeptide. The preparation can further include an adjuvant, such as Freund's complete or incomplete adjuvant, or a similar immunostimulatory agent.

Accordingly, another aspect of the invention pertains to antibodies directed against a polypeptide of the invention. The terms "antibody" and "antibody substance" as used interchangeably herein refer to immunoglobulin molecules and immunologically active portions of immunoglobulin molecules, i.e., molecules that contain an antigen binding site which specifically binds an antigen, such as a polypeptide of the invention (e.g., an epitope of a polypeptide of the invention). A molecule which specifically binds to a given polypeptide of the invention is a molecule which binds the polypeptide, but does not substantially bind other molecules in a sample, e.g., a biological sample, which naturally contains the polypeptide. Examples of immunologically active portions of immunoglobulin molecules include F(ab) and F(ab')₂ fragments which can be generated by treating the antibody with an enzyme such as pepsin. The invention provides polyclonal and monoclonal antibodies. The term "monoclonal antibody" or "monoclonal antibody composition", as used herein, refers to a population of antibody molecules that contain only one species of an antigen binding site capable of immunoreacting with a particular epitope.

Polyclonal antibodies can be prepared as described above by immunizing a suitable subject with a polypeptide of the invention as an immunogen. Preferred polyclonal antibody compositions are ones that have been selected for antibodies directed against (i.e., which bind specifically with) one or more polypeptides of the invention. Particularly preferred polyclonal antibody

PCT/US00/14858 WO 00/77239

At an appropriate time after immunization, e.g., when the specific antibody titers are highest, antibody-producing cells can be obtained from the subject and used to prepare monoclonal antibodies by standard techniques, such as the hybridoma technique originally described by Kohler and Milstein (1975) Nature 256:495-497, the human B cell hybridoma technique (Kozbor et al. (1983) Immunol. Today 4:72), the EBV-hybridoma technique (Cole et al. (1985), Monoclonal Antibodies and Cancer Therapy, Alan R. Liss, Inc., pp. 77-96) or trioma techniques. The technology for producing hybridomas is well known (see generally Current Protocols in Immunology (1994) Coligan et al. (eds.) John Wiley & Sons, Inc., New York, NY). Hybridoma cells producing a monoclonal antibody 10 of the invention are detected by screening the hybridoma culture supernatants for antibodies that bind the polypeptide of interest, e.g., using a standard ELISA assay.

5

25

30

Alternative to preparing monoclonal antibody-secreting hybridomas, a monoclonal antibody directed against a polypeptide of the invention can be identified and isolated by screening a recombinant combinatorial immunoglobulin 15 library (e.g., an antibody phage display library) with the polypeptide of interest. Kits for generating and screening phage display libraries are commercially available (e.g., the Pharmacia Recombinant Phage Antibody System, Catalog No. 27-9400-01; and the Stratagene SURFZAPTM Phage Display Kit, Catalog No. 240612). Additionally, examples of methods and reagents particularly amenable for use in 20

generating and screening antibody display library can be found in, for example, U.S. Patent No. 5,223,409; PCT Publication No. WO 92/18619; PCT Publication No. WO 91/17271; PCT Publication No. WO 92/20791; PCT Publication No. WO 92/15679; PCT Publication No. WO 93/01288; PCT Publication No. WO 92/01047; PCT Publication No. WO 92/09690; PCT Publication No. WO 90/02809; Fuchs et al. (1991) Bio/Technology 9:1370-1372; Hay et al. (1992) Hum. Antibod. Hybridomas 3:81-85; Huse et al. (1989) Science 246:1275-1281; Griffiths

Additionally, recombinant antibodies, such as chimeric and humanized monoclonal antibodies, comprising both human and non-human portions, which can be made using standard recombinant DNA techniques, are within the scope of the invention. A chimeric antibody is a molecule in which

et al. (1993) EMBO J. 12:725-734.

5

10

15

20

25

30

different portions of the antibody amino acid sequence are derived from different animal species, such as those having a variable region derived from a murine monoclonal antibody and a constant region derived from a human immunoglobulin. (See, e.g., Cabilly et al., U.S. Patent No. 4,816,567; and Boss et al., U.S. Patent No. 4,816,397). Humanized antibodies are antibody molecules which are obtained from non-human species, which have one or more complementarity-determining regions (CDRs) derived from the non-human species, and which have a framework region derived from a human immunoglobulin molecule. (See, e.g., Queen, U.S. Patent No. 5,585,089). Such chimeric and humanized monoclonal antibodies can be produced by recombinant DNA techniques known in the art, for example using methods described in PCT Publication No. WO 87/02671; European Patent Application 184,187; European Patent Application 171,496; European Patent Application 173,494; PCT Publication No. WO 86/01533; U.S. Patent No. 4,816,567; European Patent Application 125,023; Better et al. (1988) Science 240:1041-1043; Liu et al. (1987) Proc. Natl. Acad. Sci. USA 84:3439-3443; Liu et al. (1987) J. Immunol. 139:3521-3526; Sun et al. (1987) Proc. Natl. Acad. Sci. USA 84:214-218; Nishimura et al. (1987) Canc. Res. 47:999-1005; Wood et al. (1985) Nature 314:446-449; and Shaw et al. (1988) J. Natl. Cancer Inst. 80:1553-1559); Morrison (1985) Science 229:1202-1207; Oi et al. (1986) Bio/Techniques 4:214; U.S. Patent 5,225,539; Jones et al. (1986) Nature 321:552-525; Verhoeyan et al. (1988) Science 239:1534; and Beidler et al. (1988) J. Immunol. 141:4053-4060.

Completely human antibodies are particularly desirable for therapeutic treatment of human patients. Such antibodies can be produced, for example, using transgenic mice which are incapable of expressing endogenous immunoglobulin heavy and light chains genes, but which can express human heavy and light chain genes. The transgenic mice are immunized in the normal fashion with a selected antigen, e.g., all or a portion of a polypeptide of the invention. Monoclonal antibodies directed against the antigen can be obtained using conventional hybridoma technology. The human immunoglobulin transgenes harbored by the transgenic mice rearrange during B cell differentiation, and subsequently undergo class switching and somatic mutation. Thus, using such a technique, it is possible to produce therapeutically useful IgG, IgA and IgE

antibodies. For an overview of this technology for producing human antibodies, see Lonberg and Huszar (1995, *Int. Rev. Immunol.* 13:65-93). For a detailed discussion of this technology for producing human antibodies and human monoclonal antibodies and protocols for producing such antibodies, *see*, *e.g.*, U.S. Patent 5,625,126; U.S. Patent 5,633,425; U.S. Patent 5,569,825; U.S. Patent 5,661,016; and U.S. Patent 5,545,806. In addition, companies such as Abgenix, Inc. (Freemont, CA), can be engaged to provide human antibodies directed against a selected antigen using technology similar to that described above.

5

10

15

20

25

30

Completely human antibodies which recognize a selected epitope can be generated using a technique referred to as "guided selection." In this approach a selected non-human monoclonal antibody, e.g., a murine antibody, is used to guide the selection of a completely human antibody recognizing the same epitope (Jespers et al. (1994) *Bio/technology* 12:899-903).

An antibody directed against a polypeptide of the invention (e.g., monoclonal antibody) can be used to isolate the polypeptide by standard techniques, such as affinity chromatography or immunoprecipitation. Moreover, such an antibody can be used to detect the protein (e.g., in a cellular lysate or cell supernatant) in order to evaluate the abundance and pattern of expression of the polypeptide. The antibodies can also be used diagnostically to monitor protein levels in tissue as part of a clinical testing procedure, e.g., to, for example, determine the efficacy of a given treatment regimen. Detection can be facilitated by coupling the antibody to a detectable substance. Examples of detectable substances include various enzymes, prosthetic groups, fluorescent materials, luminescent materials, bioluminescent materials, and radioactive materials. Examples of suitable enzymes include horseradish peroxidase, alkaline phosphatase, βgalactosidase, or acetylcholinesterase; examples of suitable prosthetic group complexes include streptavidin/biotin and avidin/biotin; examples of suitable fluorescent materials include umbelliferone, fluorescein, fluorescein isothiocyanate, rhodamine, dichlorotriazinylamine fluorescein, dansyl chloride or phycoerythrin; an example of a luminescent material includes luminol; examples of bioluminescent materials include luciferase, luciferin, and aequorin, and examples of suitable radioactive material include 125I, 131I, 35S or 3H.

Further, an antibody substance can be conjugated with a therapeutic moiety such as a cytotoxin, a therapeutic agent, or a radioactive metal ion. Cytotoxins and cytotoxic agents include any agent that is detrimental to cells. Examples include taxol, cytochalasin B, gramicidin D, ethidium bromide, emetine, mitomycin, etoposide, tenoposide, vincristine, vinblastine, colchicin, doxorubicin, 5 daunorubicin, dihydroxy anthracin dione, mitoxantrone, mithramycin, actinomycin D, 1-dehydrotestosterone, glucocorticoids, procaine, tetracaine, lidocaine, propranolol, puromycin, and analogs or homologs of these compounds. Therapeutic agents include, antimetabolites (e.g., methotrexate, 6-mercaptopurine, 6-thioguanine, cytarabine, 5-fluorouracil, and decarbazine), alkylating agents (e.g., 10 mechlorethamine, thioepa chlorambucil, melphalan, carmustine {BSNU}, lomustine {CCNU}, cyclothosphamide, busulfan, dibromomannitol, streptozotocin, mitomycin C, and cis-dichlorodiamine platinum (II) (DDP) cisplatin), anthracyclines (e.g., daunorubicin {formerly daunomycin} and doxorubicin), antibiotics (e.g., dactinomycin {formerly actinomycin}, bleomycin, mithramycin, 15 and anthramycin {AMC}), and anti-mitotic agents (e.g., vincristine and vinblastine).

The conjugates of the invention can be used to modify a biological response; the drug moiety is not to be construed as limited to classical chemical therapeutic agents. For example, the drug moiety can be a protein or polypeptide which exhibits a desired biological activity. Such proteins include, for example, toxins such as abrin, ricin A, pseudomonas exotoxin, or diphtheria toxin; proteins such as tumor necrosis factor, alpha-interferon, beta-interferon, nerve growth factor, platelet derived growth factor, tissue plasminogen activator; and biological response modifiers such as lymphokines, interleukin-1 (IL-1), interleukin-2 (IL-2), interleukin-6 (IL-6), granulocyte macrophage colony stimulating factor (GM-CSF), granulocyte colony stimulating factor (G-CSF), and other growth factors.

20

25

30

Techniques for conjugating a therapeutic moiety with an antibody substance are well known (see, e.g., Arnon et al., "Monoclonal Antibodies For Immunotargeting Of Drugs In Cancer Therapy", in Monoclonal Antibodies and Cancer Therapy, Reisfeld et al., eds., pp. 243-256, Alan R. Liss, Inc., 1985; Hellstrom et al., "Antibodies For Drug Delivery", in Controlled Drug Delivery, 2nd

Ed., Robinson et al., eds., pp. 623-653, Marcel Dekker, Inc., 1987; Thorpe,

"Antibody Carriers Of Cytotoxic Agents In Cancer Therapy: A Review", in

Monoclonal Antibodies '84: Biological and Clinical Applications, Pinchera et al.,
eds., pp. 475-506, 1985; "Analysis, Results, And Future Prospective Of The

Therapeutic Use Of Radiolabeled Antibody In Cancer Therapy", in Monoclonal
Antibodies for Cancer Detection and Therapy, Baldwin et al., eds., pp. 303-316,
Academic Press, 1985; and Thorpe et al., "The Preparation And Cytotoxic
Properties Of Antibody-Toxin Conjugates", Immunol. Rev. 62:119-58, 1982).
Alternatively, an antibody can be conjugated with a second antibody to form an
antibody heteroconjugate as described by Segal in U.S. Patent No. 4,676,980.

Accordingly, in one aspect, the invention provides substantially purified antibodies or fragment thereof, and non-human antibodies or fragments thereof, which antibodies or fragments specifically bind with a polypeptide having an amino acid sequence which comprises a sequence selected from the group consisting of:

15

20

25

30

- (i) SEQ ID NOs: 3-8, 11-16, 19-24, 27-32, 35-44, 47-52, 55-66, 69, and 74;
- (ii) the amino acid sequence encoded by a cDNA of a clone deposited as one of ATCC[®] 207219, 207184, 207228, 207185, 207220, and 207221;
- (iii) at least 15 amino acid residues of the amino acid sequence of any of SEQ ID NOs: 3-8, 11-16, 19-24, 27-32, 35-44, 47-52, 55-66, 69, and 74;
 - (iv) an amino acid sequence which is at least 95% identical to the amino acid sequence of any of SEQ ID NOs: 3-8, 11-16, 19-24, 27-32, 35-44, 47-52, 55-66, 69, and 74, wherein the percent identity is determined using the ALIGN program of the GCG software package with a PAM120 weight residue table, a gap length penalty of 12, and a gap penalty of 4; and
 - (v) an amino acid sequence which is encoded by a nucleic acid molecule which hybridizes with a nucleic acid having a sequence selected from the group consisting of SEQ ID NOs: 1, 2, 9, 10, 17, 18, 25, 26, 33, 34, 45, 46, 53, 54, 67, 68, 72, and 73 under conditions of hybridization of 6 × SSC (standard saline citrate) at 45°C and washing in 0.2 × SSC, 0.1% SDS at 65°C.

In another aspect, the invention provides non-human antibodies or fragments thereof, which antibodies or fragments specifically bind with a

polypeptide having an amino acid sequence which comprises a sequence selected from the group consisting of:

- (i) SEQ ID NOs: 3-8, 11-16, 19-24, 27-32, 35-44, 47-52, 55-66, 69, and 74;
- (ii) the amino acid sequence encoded by a cDNA of a clone deposited as one of ATCC[®] 207219, 207184, 207228, 207185, 207220, and 207221;
- (iii) at least 15 amino acid residues of the amino acid sequence of any of SEO ID NOs: 3-8, 11-16, 19-24, 27-32, 35-44, 47-52, 55-66, 69, and 74;

5

10

15

20

25

- (iv) an amino acid sequence which is at least 95% identical to the amino acid sequence of any of SEQ ID NOs: 3-8, 11-16, 19-24, 27-32, 35-44, 47-52, 55-66, 69, and 74, wherein the percent identity is determined using the ALIGN program of the GCG software package with a PAM120 weight residue table, a gap length penalty of 12, and a gap penalty of 4; and
- (v) an amino acid sequence which is encoded by a nucleic acid molecule which hybridizes with a nucleic acid having a sequence selected from the group consisting of SEQ ID NOs: 1, 2, 9, 10, 17, 18, 25, 26, 33, 34, 45, 46, 53, 54, 67, 68, 72, and 73 under conditions of hybridization of 6 × SSC (standard saline citrate) at 45°C and washing in 0.2 × SSC, 0.1% SDS at 65°C. Such non-human antibodies can be goat, mouse, sheep, horse, chicken, rabbit, or rat antibodies. Alternatively, the non-human antibodies of the invention can be chimeric and/or humanized antibodies. In addition, the non-human antibodies of the invention can be polyclonal antibodies or monoclonal antibodies.

In still a further aspect, the invention provides monoclonal antibodies or fragments thereof, which antibodies or fragments specifically bind with a polypeptide having an amino acid sequence which comprises a sequence selected from the group consisting of:

- (i) SEQ ID NOs: 3-8, 11-16, 19-24, 27-32, 35-44, 47-52, 55-66, 69, and 74;
- (ii) the amino acid sequence encoded by a cDNA of a clone deposited as one of ATCC[®] 207219, 207184, 207228, 207185, 207220, and 207221;
- (iii) at least 15 amino acid residues of the amino acid sequence of any of SEQ ID NOs: 3-8, 11-16, 19-24, 27-32, 35-44, 47-52, 55-66, 69, and 74;
 - (iv) an amino acid sequence which is at least 95% identical to the amino acid sequence of any of SEQ ID NOs: 3-8, 11-16, 19-24, 27-32, 35-44, 47-52, 55-

66, 69, and 74, wherein the percent identity is determined using the ALIGN program of the GCG software package with a PAM120 weight residue table, a gap length penalty of 12, and a gap penalty of 4; and

(v) an amino acid sequence which is encoded by a nucleic acid molecule which hybridizes with a nucleic acid having a sequence selected from the group consisting of SEQ ID NOs: 1, 2, 9, 10, 17, 18, 25, 26, 33, 34, 45, 46, 53, 54, 67, 68, 72, and 73 under conditions of hybridization of 6 × SSC (standard saline citrate) at 45°C and washing in 0.2 × SSC, 0.1% SDS at 65°C. The monoclonal antibodies can be human, humanized, chimeric and/or non-human antibodies.

5

10

15

20

25

30

The substantially purified antibodies or fragments thereof can specifically bind with a signal peptide, a secreted sequence, an extracellular domain, a transmembrane or a cytoplasmic domain cytoplasmic membrane of a polypeptide of the invention. In a particularly preferred embodiment, the substantially purified antibodies or fragments thereof, the non-human antibodies or fragments thereof, and/or the monoclonal antibodies or fragments thereof, of the invention specifically bind with a secreted sequence or with an extracellular domain of one of TANGO 202, TANGO 234, TANGO 265, TANGO 273, TANGO 286, TANGO 294, and INTERCEPT 296. Preferably, the extracellular domain with which the antibody substance binds has an amino acid sequence selected from the group consisting of SEQ ID NOs: 5, 6, 14, 22, 30, 37, 49, 50, and 56-58.

Any of the antibody substances of the invention can be conjugated with a therapeutic moiety or to a detectable substance. Non-limiting examples of detectable substances that can be conjugated with the antibody substances of the invention include an enzyme, a prosthetic group, a fluorescent material (i.e., a fluorophore), a luminescent material, a bioluminescent material, and a radioactive material (e.g., a radionuclide or a substituent comprising a radionuclide)..

The invention also provides a kit containing an antibody substance of the invention conjugated with a detectable substance, and instructions for use. Still another aspect of the invention is a pharmaceutical composition comprising an antibody substance of the invention and a pharmaceutically acceptable carrier. In preferred embodiments, the pharmaceutical composition contains an antibody

substance of the invention, a therapeutic moiety (preferably conjugated with the antibody substance), and a pharmaceutically acceptable carrier.

Still another aspect of the invention is a method of making an antibody that specifically recognizes one of TANGO 202, TANGO 234, TANGO 265, TANGO 273, TANGO 286, TANGO 294, and INTERCEPT 296. This method comprises immunizing a vertebrate (e.g., a mammal such as a rabbit, goat, or pig) with a polypeptide. The polypeptide used as an immunogen has an amino acid sequence that comprises a sequence selected from the group consisting of:

5

20

25

30

- (i) SEQ ID NOs: 3-8, 11-16, 19-24, 27-32, 35-44, 47-52, 55-66, 69, and 74;
- (ii) the amino acid sequence encoded by a cDNA of a clone deposited as one of ATCC[®] 207219, 207184, 207228, 207185, 207220, and 207221;
 - (iii) at least 15 amino acid residues of the amino acid sequence of any of SEO ID NOs: 3-8, 11-16, 19-24, 27-32, 35-44, 47-52, 55-66, 69, and 74;
- (iv) an amino acid sequence which is at least 95% identical to the amino
 acid sequence of any of SEQ ID NOs: 3-8, 11-16, 19-24, 27-32, 35-44, 47-52, 5566, 69, and 74, wherein the percent identity is determined using the ALIGN
 program of the GCG software package with a PAM120 weight residue table, a gap
 length penalty of 12, and a gap penalty of 4; and
 - (v) an amino acid sequence which is encoded by a nucleic acid molecule which hybridizes with a nucleic acid having a sequence selected from the group consisting of SEQ ID NOs: 1, 2, 9, 10, 17, 18, 25, 26, 33, 34, 45, 46, 53, 54, 67, 68, 72, and 73 under conditions of hybridization of 6 × SSC (standard saline citrate) at 45°C and washing in 0.2 × SSC, 0.1% SDS at 65°C.

After immunization, a sample is collected from the vertebrate that contains an antibody that specifically recognizes the polypeptide with which the vertebrate was immunized. Preferably, the polypeptide is recombinantly produced using a non-human host cell. Optionally, an antibody substance can be further purified from the sample using techniques well known to those of skill in the art. The method can further comprise making a monoclonal antibody-producing cell from a cell of the vertebrate. Optionally, antibodies can be collected from the antibody-producing cell.

III. Recombinant Expression Vectors and Host Cells

5

10

15

20

25

30

Another aspect of the invention pertains to vectors, preferably expression vectors, containing a nucleic acid encoding a polypeptide of the invention (or a portion thereof). As used herein, the term "vector" refers to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. One type of vector is a "plasmid", which refers to a circular double stranded DNA loop into which additional DNA segments can be ligated. Another type of vector is a viral vector, wherein additional DNA segments can be ligated into the viral genome. Certain vectors are capable of autonomous replication in a host cell into which they are introduced (e.g., bacterial vectors having a bacterial origin of replication and episomal mammalian vectors). Other vectors (e.g., nonepisomal mammalian vectors) are integrated into the genome of a host cell upon introduction into the host cell, and thereby are replicated along with the host genome. Moreover, certain vectors, expression vectors, are capable of directing the expression of genes to which they are operably linked. In general, expression vectors of utility in recombinant DNA techniques are often in the form of plasmids (vectors). However, the invention is intended to include such other forms of expression vectors, such as viral vectors (e.g., replication defective retroviruses, adenoviruses and adeno-associated viruses), which serve equivalent functions.

The recombinant expression vectors of the invention comprise a nucleic acid of the invention in a form suitable for expression of the nucleic acid in a host cell. This means that the recombinant expression vectors include one or more regulatory sequences, selected on the basis of the host cells to be used for expression, which is operably linked to the nucleic acid sequence to be expressed. Within a recombinant expression vector, "operably linked" is intended to mean that the nucleotide sequence of interest is linked to the regulatory sequence(s) in a manner which allows for expression of the nucleotide sequence (e.g., in an *in vitro* transcription/translation system or in a host cell when the vector is introduced into the host cell). The term "regulatory sequence" is intended to include promoters, enhancers and other expression control elements (e.g., polyadenylation signals). Such regulatory sequences are described, for example, in Goeddel, *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, CA (1990).

Regulatory sequences include those which direct constitutive expression of a nucleotide sequence in many types of host cell and those which direct expression of the nucleotide sequence only in certain host cells (e.g., tissue-specific regulatory sequences). It will be appreciated by those skilled in the art that the design of the expression vector can depend on such factors as the choice of the host cell to be transformed, the level of expression of protein desired, and the like. The expression vectors of the invention can be introduced into host cells to thereby produce proteins or peptides, including fusion proteins or peptides, encoded by nucleic acids as described herein.

5

10

15

20

25

30

The recombinant expression vectors of the invention can be designed for expression of a polypeptide of the invention in prokaryotic (e.g., *E. coli*) or eukaryotic cells (e.g., insect cells (using baculovirus expression vectors), yeast cells or mammalian cells). Suitable host cells are discussed further in Goeddel, *supra*. Alternatively, the recombinant expression vector can be transcribed and translated *in vitro*, for example using T7 promoter regulatory sequences and T7 polymerase.

Expression of proteins in prokaryotes is most often carried out in E. coli with vectors containing constitutive or inducible promoters directing the expression of either fusion or non-fusion proteins. Fusion vectors add a number of amino acids to a protein encoded therein, usually to the amino terminus of the recombinant protein. Such fusion vectors typically serve three purposes: 1) to increase expression of recombinant protein; 2) to increase the solubility of the recombinant protein; and 3) to aid in the purification of the recombinant protein by acting as a ligand in affinity purification. Often, in fusion expression vectors, a proteolytic cleavage site is introduced at the junction of the fusion moiety and the recombinant protein to enable separation of the recombinant protein from the fusion moiety subsequent to purification of the fusion protein. Such enzymes, and their cognate recognition sequences, include Factor Xa, thrombin and enterokinase. Typical fusion expression vectors include pGEX (Pharmacia Biotech Inc; Smith and Johnson (1988) Gene 67:31-40), pMAL (New England Biolabs, Beverly, MA) and pRIT5 (Pharmacia, Piscataway, NJ) which fuse glutathione S-transferase (GST), maltose E binding protein, or protein A, respectively, to the target recombinant protein.

Examples of suitable inducible non-fusion *E. coli* expression vectors include pTrc (Amann et al., (1988) *Gene* 69:301-315) and pET 11d (Studier et al., *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, California (1990) 60-89). Target gene expression from the pTrc vector relies on host RNA polymerase transcription from a hybrid trp-lac fusion promoter. Target gene expression from the pET 11d vector relies on transcription from a T7 gn10-lac fusion promoter mediated by a co-expressed viral RNA polymerase (T7 gn1). This viral polymerase is supplied by host strains BL21(DE3) or HMS174(DE3) from a resident lambda prophage harboring a T7 gn1 gene under the transcriptional control of the lacUV 5 promoter.

5

10

15

30

One strategy to maximize recombinant protein expression in *E. coli* is to express the protein in a host bacteria with an impaired capacity to proteolytically cleave the recombinant protein (Gottesman, *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, California (1990) 119-128). Another strategy is to alter the nucleic acid sequence of the nucleic acid to be inserted into an expression vector so that the individual codons for each amino acid are those preferentially utilized in *E. coli* (Wada et al. (1992) *Nucleic Acids Res.* 20:2111-2118). Such alteration of nucleic acid sequences of the invention can be carried out by standard DNA synthesis techniques.

In another embodiment, the expression vector is a yeast expression vector. Examples of vectors for expression in yeast *S. cerevisiae* include pYepSec1 (Baldari et al. (1987) *EMBO J.* 6:229-234), pMFa (Kurjan and Herskowitz, (1982) *Cell* 30:933-943), pJRY88 (Schultz et al. (1987) *Gene* 54:113-123), pYES2 (Invitrogen Corporation, San Diego, CA), and pPicZ (Invitrogen Corp, San Diego, CA).

Alternatively, the expression vector is a baculovirus expression vector. Baculovirus vectors available for expression of proteins in cultured insect cells (e.g., Sf 9 cells) include the pAc series (Smith et al. (1983) *Mol. Cell Biol.* 3:2156-2165) and the pVL series (Lucklow and Summers (1989) *Virology* 170:31-39).

In yet another embodiment, a nucleic acid of the invention is expressed in mammalian cells using a mammalian expression vector. Examples of

mammalian expression vectors include pCDM8 (Seed (1987) Nature 329:840) and pMT2PC (Kaufman et al. (1987) EMBO J. 6:187-195). When used in mammalian cells, the expression vector's control functions are often provided by viral regulatory elements. For example, commonly used promoters are derived from polyoma, Adenovirus 2, cytomegalovirus and Simian Virus 40. For other suitable expression systems for both prokaryotic and eukaryotic cells see chapters 16 and 17 of Sambrook et al., supra.

10

15

20

25

30

In another embodiment, the recombinant mammalian expression vector is capable of directing expression of the nucleic acid preferentially in a particular cell type (e.g., tissue-specific regulatory elements are used to express the nucleic acid). Tissue-specific regulatory elements are known in the art. Nonlimiting examples of suitable tissue-specific promoters include the albumin promoter (liver-specific; Pinkert et al. (1987) Genes Dev. 1:268-277), lymphoidspecific promoters (Calame and Eaton (1988) Adv. Immunol. 43:235-275), in particular promoters of T cell receptors (Winoto and Baltimore (1989) EMBO J. 8:729-733) and immunoglobulins (Banerji et al. (1983) Cell 33:729-740; Queen and Baltimore (1983) Cell 33:741-748), neuron-specific promoters (e.g., the neurofilament promoter; Byrne and Ruddle (1989) Proc. Natl. Acad. Sci. USA 86:5473-5477), pancreas-specific promoters (Edlund et al. (1985) Science 230:912-916), and mammary gland-specific promoters (e.g., milk whey promoter; U.S. Patent No. 4,873,316 and European Application Publication No. 264,166). Developmentally-regulated promoters are also encompassed, for example the murine hox promoters (Kessel and Gruss (1990) Science 249:374-379) and the αfetoprotein promoter (Campes and Tilghman (1989) Genes Dev. 3:537-546).

The invention further provides a recombinant expression vector comprising a DNA molecule of the invention cloned into the expression vector in an antisense orientation. That is, the DNA molecule is operably linked to a regulatory sequence in a manner which allows for expression (by transcription of the DNA molecule) of an RNA molecule which is antisense to the mRNA encoding a polypeptide of the invention. Regulatory sequences operably linked to a nucleic acid cloned in the antisense orientation can be chosen which direct the continuous expression of the antisense RNA molecule in a variety of cell types, for instance

viral promoters and/or enhancers, or regulatory sequences can be chosen which direct constitutive, tissue specific or cell type specific expression of antisense RNA. The antisense expression vector can be in the form of a recombinant plasmid, phagemid, or attenuated virus in which antisense nucleic acids are produced under the control of a high efficiency regulatory region, the activity of which can be determined by the cell type into which the vector is introduced. For a discussion of the regulation of gene expression using antisense genes see Weintraub et al. (Reviews - Trends in Genetics, Vol. 1(1) 1986).

5

10

15

20

25

30

Another aspect of the invention pertains to host cells into which a recombinant expression vector of the invention has been introduced. The terms "host cell" and "recombinant host cell" are used interchangeably herein. It is understood that such terms refer not only to the particular subject cell but to the progeny or potential progeny of such a cell. Because certain modifications may occur in succeeding generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term as used herein.

A host cell can be any prokaryotic (e.g., E. coli) or eukaryotic cell (e.g., insect cells, yeast or mammalian cells).

Vector DNA can be introduced into prokaryotic or eukaryotic cells via conventional transformation or transfection techniques. As used herein, the terms "transformation" and "transfection" are intended to refer to a variety of art-recognized techniques for introducing foreign nucleic acid into a host cell, including calcium phosphate or calcium chloride co-precipitation, DEAE-dextran-mediated transfection, lipofection, or electroporation. Suitable methods for transforming or transfecting host cells can be found in Sambrook, et al. (*supra*), and other laboratory manuals.

For stable transfection of mammalian cells, it is known that, depending upon the expression vector and transfection technique used, only a small fraction of cells may integrate the foreign DNA into their genome. In order to identify and select these integrants, a gene that encodes a selectable marker (e.g., for resistance to antibiotics) is generally introduced into the host cells along with the gene of interest. Preferred selectable markers include those which confer resistance

to drugs, such as G418, hygromycin and methotrexate. Cells stably transfected with the introduced nucleic acid can be identified by drug selection (e.g., cells that have incorporated the selectable marker gene will survive, while the other cells die).

In another embodiment, the expression characteristics of an endogenous nucleic acid within a cell, cell line, or microorganism (e.g., a TANGO 5 202, TANGO 234, TANGO 265, TANGO 273, TANGO 286, TANGO 294, or INTERCEPT 296 nucleic acid, as described herein) can be modified by inserting a heterologous DNA regulatory element (i.e., one that is heterologous with respect to the endogenous gene) into the genome of the cell, stable cell line, or cloned microorganism. The inserted regulatory element can be operatively linked with the 10 endogenous gene (e.g., TANGO 202, TANGO 234, TANGO 265, TANGO 273, TANGO 286, TANGO 294, or INTERCEPT 296) and thereby control, modulate, or activate the endogenous gene. For example, an endogenous TANGO 202, TANGO 234, TANGO 265, TANGO 273, TANGO 286, TANGO 294, or INTERCEPT 296 gene which is normally "transcriptionally silent" (i.e., a TANGO 202, TANGO 234, 15 TANGO 265, TANGO 273, TANGO 286, TANGO 294, or INTERCEPT 296 gene which is normally not expressed, or is normally expressed only at only a very low level) can be activated by inserting a regulatory element which is capable of promoting expression of the gene in the cell, cell line, or microorganism. Alternatively, a transcriptionally silent, endogenous TANGO 202, TANGO 234, 20 TANGO 265, TANGO 273, TANGO 286, TANGO 294, or INTERCEPT 296 gene can be activated by inserting a promiscuous regulatory element that works across cell types.

A heterologous regulatory element can be inserted into a stable cell
line or cloned microorganism such that it is operatively linked with and activates expression of an endogenous TANGO 202, TANGO 234, TANGO 265, TANGO 273, TANGO 286, TANGO 294, or INTERCEPT 296 gene, using techniques, such as targeted homologous recombination, which are well known to those of skill in the art (described e.g., in Chappel, U.S. Patent No. 5,272,071; PCT publication No. WO 91/06667, published May 16, 1991).

A host cell of the invention, such as a prokaryotic or eukaryotic host cell in culture, can be used to produce a polypeptide of the invention. Accordingly,

the invention further provides methods for producing a polypeptide of the invention using the host cells of the invention. In one embodiment, the method comprises culturing the host cell of invention (into which a recombinant expression vector encoding a polypeptide of the invention has been introduced) in a suitable medium such that the polypeptide is produced. In another embodiment, the method further comprises isolating the polypeptide from the medium or the host cell.

5

10

15

20

25

The host cells of the invention can also be used to produce nonhuman transgenic animals. For example, in one embodiment, a host cell of the invention is a fertilized oocyte or an embryonic stem cell into which a sequences encoding a polypeptide of the invention have been introduced. Such host cells can then be used to create non-human transgenic animals in which exogenous sequences encoding a polypeptide of the invention have been introduced into their genome or homologous recombinant animals in which endogenous encoding a polypeptide of the invention sequences have been altered. Such animals are useful for studying the function and/or activity of the polypeptide and for identifying and/or evaluating modulators of polypeptide activity. As used herein, a "transgenic animal" is a nonhuman animal, preferably a mammal, more preferably a rodent such as a rat or mouse, in which one or more of the cells of the animal includes a transgene. Other examples of transgenic animals include non-human primates, sheep, dogs, cows, goats, chickens, amphibians, etc. A transgene is exogenous DNA which is integrated into the genome of a cell from which a transgenic animal develops and which remains in the genome of the mature animal, thereby directing the expression of an encoded gene product in one or more cell types or tissues of the transgenic animal. As used herein, an "homologous recombinant animal" is a non-human animal, preferably a mammal, more preferably a mouse, in which an endogenous gene has been altered by homologous recombination between the endogenous gene and an exogenous DNA molecule introduced into a cell of the animal, e.g., an embryonic cell of the animal, prior to development of the animal.

A transgenic animal of the invention can be created by introducing
nucleic acid encoding a polypeptide of the invention (or a homologue thereof) into
the male pronuclei of a fertilized oocyte, e.g., by microinjection, retroviral
infection, and allowing the oocyte to develop in a pseudopregnant female foster

animal. Intronic sequences and polyadenylation signals can also be included in the transgene to increase the efficiency of expression of the transgene. A tissue-specific regulatory sequence(s) can be operably linked to the transgene to direct expression of the polypeptide of the invention to particular cells. Methods for generating transgenic animals via embryo manipulation and microinjection, particularly animals such as mice, have become conventional in the art and are described, for example, in U.S. Patent Nos. 4,736,866 and 4,870,009, U.S. Patent No. 4,873,191, in Hogan, Manipulating the Mouse Embryo, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1986, and in Wakayama et al., 1999, Proc. Natl. Acad. Sci. USA 96:14984-14989. Similar methods can be used to produce other 10 transgenic animals. A transgenic founder animal can be identified based upon the presence of the transgene in its genome and/or expression of mRNA encoding the transgene in tissues or cells of the animals. A transgenic founder animal can then be used to breed additional animals carrying the transgene. Moreover, transgenic animals carrying the transgene can further be bred to other transgenic animals 15 carrying other transgenes.

To create a homologous recombinant animal, a vector is prepared which contains at least a portion of a gene encoding a polypeptide of the invention into which a deletion, addition or substitution has been introduced to thereby alter, e.g., functionally disrupt, the gene. In a preferred embodiment, the vector is designed such that, upon homologous recombination, the endogenous gene is functionally disrupted (i.e., no longer encodes a functional protein; also referred to as a "knock out" vector). Alternatively, the vector can be designed such that, upon homologous recombination, the endogenous gene is mutated or otherwise altered but still encodes functional protein (e.g., the upstream regulatory region can be altered to thereby alter the expression of the endogenous protein). In the homologous recombination vector, the altered portion of the gene is flanked at its 5' and 3' ends by additional nucleic acid of the gene to allow for homologous recombination to occur between the exogenous gene carried by the vector and an endogenous gene in an embryonic stem cell. The additional flanking nucleic acid sequences are of sufficient length for successful homologous recombination with the endogenous gene. Typically, several kilobases of flanking DNA (both at the 5'

20

25

30

and 3' ends) are included in the vector (see, e.g., Thomas and Capecchi (1987) Cell 51:503 for a description of homologous recombination vectors). The vector is introduced into an embryonic stem cell line (e.g., by electroporation) and cells in which the introduced gene has homologously recombined with the endogenous gene are selected (see, e.g., Li et al. (1992) Cell 69:915). The selected cells are then injected into a blastocyst of an animal (e.g., a mouse) to form aggregation chimeras (see, e.g., Bradley in Teratocarcinomas and Embryonic Stem Cells: A Practical Approach, Robertson, ed. (IRL, Oxford, 1987) pp. 113-152). A chimeric embryo can then be implanted into a suitable pseudopregnant female foster animal and the embryo brought to term. Progeny harboring the homologously recombined DNA in their germ cells can be used to breed animals in which all cells of the animal contain the homologously recombined DNA by germline transmission of the transgene. Methods for constructing homologous recombination vectors and homologous recombinant animals are described further in Bradley (1991) Current Opinion in Bio/Technology 2:823-829 and in PCT Publication Nos. WO 90/11354, WO 91/01140, WO 92/0968, and WO 93/04169.

5

10

15

20

25

In another embodiment, transgenic non-human animals can be produced which contain selected systems which allow for regulated expression of the transgene. One example of such a system is the *cre/loxP* recombinase system of bacteriophage P1. For a description of the *cre/loxP* recombinase system, *see*, *e.g.*, Lakso et al. (1992) *Proc. Natl. Acad. Sci. USA* 89:6232-6236. Another example of a recombinase system is the FLP recombinase system of *Saccharomyces cerevisiae* (O'Gorman et al. (1991) *Science* 251:1351-1355. If a *cre/loxP* recombinase system is used to regulate expression of the transgene, animals containing transgenes encoding both the *Cre* recombinase and a selected protein are required. Such animals can be provided through the construction of "double" transgenic animals, e.g., by mating two transgenic animals, one containing a transgene encoding a selected protein and the other containing a transgene encoding a recombinase.

Clones of the non-human transgenic animals described herein can
also be produced according to the methods described in Wilmut et al. (1997) *Nature*385:810-813 and PCT Publication Nos. WO 97/07668 and WO 97/07669.

IV. Pharmaceutical Compositions

5

10

15

20

25

30

The nucleic acid molecules, polypeptides, and antibodies (also referred to herein as "active compounds") of the invention can be incorporated into pharmaceutical compositions suitable for administration. Such compositions typically comprise the nucleic acid molecule, protein, or antibody and a pharmaceutically acceptable carrier. As used herein the language "pharmaceutically acceptable carrier" is intended to include any and all solvents, dispersion media, coatings, antibacterial and antifungal agents, isotonic and absorption delaying agents, and the like, compatible with pharmaceutical administration. The use of such media and agents for pharmaceutically active substances is well known in the art. Except insofar as any conventional media or agent is incompatible with the active compound, use thereof in the compositions is contemplated. Supplementary active compounds can also be incorporated into the compositions.

The invention includes methods for preparing pharmaceutical compositions for modulating the expression or activity of a polypeptide or nucleic acid of the invention. Such methods comprise formulating a pharmaceutically acceptable carrier with an agent which modulates expression or activity of a polypeptide or nucleic acid of the invention. Such compositions can further include additional active agents. Thus, the invention further includes methods for preparing a pharmaceutical composition by formulating a pharmaceutically acceptable carrier with an agent which modulates expression or activity of a polypeptide or nucleic acid of the invention and one or more additional active compounds.

The agent which modulates expression or activity can, for example, be a small molecule other than a nucleic acid, polypeptide, or antibody of the invention. For example, such small molecules include peptides, peptidomimetics, amino acids, amino acid analogs, polynucleotides, polynucleotide analogs, nucleotides, nucleotide analogs, organic or inorganic compounds (i.e., including heteroorganic and organometallic compounds) having a molecular weight less than about 10,000 grams per mole, organic or inorganic compounds having a molecular weight less than about 5,000 grams per mole, organic or inorganic compounds having a molecular weight less than about 1,000 grams per mole, organic or inorganic compounds having a molecular weight less than about 500 grams per

mole, and salts, esters, and other pharmaceutically acceptable forms of such compounds.

5

10

15

20

25

30

It is understood that appropriate doses of small molecule agents and protein or polypeptide agents depends upon a number of factors within the ken of the ordinarily skilled physician, veterinarian, or researcher. The dose(s) of these agents will vary, for example, depending upon the identity, size, and condition of the subject or sample being treated, further depending upon the route by which the composition is to be administered, if applicable, and the effect which the practitioner desires the agent to have upon the nucleic acid or polypeptide of the invention. Exemplary doses of a small molecule include milligram or microgram amounts per kilogram of subject or sample weight (e.g., about 1 microgram per kilogram to about 500 milligrams per kilogram, about 100 micrograms per kilogram to about 5 milligrams per kilogram, or about 1 microgram per kilogram to about 50 micrograms per kilogram). Exemplary doses of a protein or polypeptide include gram, milligram or microgram amounts per kilogram of subject or sample weight (e.g., about 1 microgram per kilogram to about 5 grams per kilogram, about 100 micrograms per kilogram to about 500 milligrams per kilogram, or about 1 milligram per kilogram to about 50 milligrams per kilogram). It is furthermore understood that appropriate doses of one of these agents depend upon the potency of the agent with respect to the expression or activity to be modulated. Such appropriate doses can be determined using the assays described herein. When one or more of these agents is to be administered to an animal (e.g., a human) in order to modulate expression or activity of a polypeptide or nucleic acid of the invention, a physician, veterinarian, or researcher can, for example, prescribe a relatively low dose at first, subsequently increasing the dose until an appropriate response is obtained. In addition, it is understood that the specific dose level for any particular animal subject will depend upon a variety of factors including the activity of the specific agent employed, the age, body weight, general health, gender, and diet of the subject, the time of administration, the route of administration, the rate of excretion, any drug combination, and the degree of expression or activity to be modulated.

A pharmaceutical composition of the invention is formulated to be compatible with its intended route of administration. Examples of routes of administration include parenteral, e.g., intravenous, intradermal, subcutaneous, oral (e.g., inhalation), transdermal (topical), transmucosal, and rectal administration. Solutions or suspensions used for parenteral, intradermal, or subcutaneous application can include the following components: a sterile diluent such as water for injection, saline solution, fixed oils, polyethylene glycols, glycerine, propylene glycol or other synthetic solvents; antibacterial agents such as benzyl alcohol or methyl parabens; antioxidants such as ascorbic acid or sodium bisulfite; chelating agents such as ethylenediamine-tetraacetic acid; buffers such as acetates, citrates or phosphates and agents for the adjustment of tonicity such as sodium chloride or dextrose. pH can be adjusted with acids or bases, such as hydrochloric acid or sodium hydroxide. The parenteral preparation can be enclosed in ampules, disposable syringes or multiple dose vials made of glass or plastic.

5

10

15

20

25

30

Pharmaceutical compositions suitable for injectable use include sterile aqueous solutions (where water soluble) or dispersions and sterile powders for the extemporaneous preparation of sterile injectable solutions or dispersions. For intravenous administration, suitable carriers include physiological saline, bacteriostatic water, Cremophor EL (BASF; Parsippany, NJ) or phosphate buffered saline (PBS). In all cases, the composition must be sterile and should be fluid to the extent that easy syringability exists. It must be stable under the conditions of manufacture and storage and must be preserved against the contaminating action of microorganisms such as bacteria and fungi. The carrier can be a solvent or dispersion medium containing, for example, water, ethanol, polyol (for example, glycerol, propylene glycol, and liquid polyethylene glycol, and the like), and suitable mixtures thereof. The proper fluidity can be maintained, for example, by the use of a coating such as lecithin, by the maintenance of the required particle size in the case of dispersion and by the use of surfactants. Prevention of the action of microorganisms can be achieved by various antibacterial and antifungal agents, for example, parabens, chlorobutanol, phenol, ascorbic acid, thimerosal, and the like. In many cases, it will be preferable to include isotonic agents, for example, sugars, polyalcohols such as mannitol, sorbitol, or sodium chloride in the composition.

Prolonged absorption of the injectable compositions can be brought about by including in the composition an agent which delays absorption, for example, aluminum monostearate and gelatin.

5

10

15

20

25

30

Sterile injectable solutions can be prepared by incorporating the active compound (e.g., a polypeptide or antibody) in the required amount in an appropriate solvent with one or a combination of ingredients enumerated above, as required, followed by filtered sterilization. Generally, dispersions are prepared by incorporating the active compound into a sterile vehicle which contains a basic dispersion medium, and then incorporating the required other ingredients from those enumerated above. In the case of sterile powders for the preparation of sterile injectable solutions, the preferred methods of preparation are vacuum drying and freeze-drying which yields a powder of the active ingredient plus any additional desired ingredient from a previously sterile-filtered solution thereof.

Oral compositions generally include an inert diluent or an edible carrier. They can be enclosed in gelatin capsules or compressed into tablets. For the purpose of oral therapeutic administration, the active compound can be incorporated with excipients and used in the form of tablets, troches, or capsules. Oral compositions can also be prepared using a fluid carrier for use as a mouthwash, wherein the compound in the fluid carrier is applied orally and swished and expectorated or swallowed.

Pharmaceutically compatible binding agents, and/or adjuvant materials can be included as part of the composition. The tablets, pills, capsules, troches, and the like can contain any of the following ingredients, or compounds of a similar nature: a binder such as microcrystalline cellulose, gum tragacanth or gelatin; an excipient such as starch or lactose, a disintegrating agent such as alginic acid, Primogel, or corn starch; a lubricant such as magnesium stearate or Sterotes; a glidant such as colloidal silicon dioxide; a sweetening agent such as sucrose or saccharin; or a flavoring agent such as peppermint, methyl salicylate, or orange flavoring.

For administration by inhalation, the compounds are delivered in the form of an aerosol spray from a pressurized container or dispenser which contains a suitable propellant, e.g., a gas such as carbon dioxide, or a nebulizer.

Systemic administration can also be by transmucosal or transdermal means. For transmucosal or transdermal administration, penetrants appropriate to the barrier to be permeated are used in the formulation. Such penetrants are generally known in the art, and include, for example, for transmucosal administration, detergents, bile salts, and fusidic acid derivatives. Transmucosal administration can be accomplished through the use of nasal sprays or suppositories. For transdermal administration, the active compounds are formulated into ointments, salves, gels, or creams as generally known in the art.

5

10

15

20

25

30

The compounds can also be prepared in the form of suppositories (e.g., with conventional suppository bases such as cocoa butter and other glycerides) or retention enemas for rectal delivery.

In one embodiment, the active compounds are prepared with carriers that will protect the compound against rapid elimination from the body, such as a controlled release formulation, including implants and microencapsulated delivery systems. Biodegradable, biocompatible polymers can be used, such as ethylene vinyl acetate, polyanhydrides, polyglycolic acid, collagen, polyorthoesters, and polylactic acid. Methods for preparation of such formulations will be apparent to those skilled in the art. The materials can also be obtained commercially from Alza Corporation and Nova Pharmaceuticals, Inc. Liposomal suspensions (including liposomes which can be targeted to bind with virus-infected cells using a monoclonal antibody which binds specifically with a viral antigen) can also be used as pharmaceutically acceptable carriers. These can be prepared according to methods known to those skilled in the art, for example, as described in U.S. Patent No. 4,522,811.

It is especially advantageous to formulate oral or parenteral compositions in dosage unit form for ease of administration and uniformity of dosage. Dosage unit form as used herein refers to physically discrete units suited as unitary dosages for the subject to be treated; each unit containing a predetermined quantity of active compound calculated to produce the desired therapeutic effect in association with the required pharmaceutical carrier. The specification for the dosage unit forms of the invention are dictated by and directly dependent on the unique characteristics of the active compound and the particular therapeutic effect

to be achieved, and the limitations inherent in the art of compounding such an active compound for the treatment of individuals.

For antibodies, the preferred dosage is 0.1 mg/kg to 100 mg/kg of body weight (generally 10 mg/kg to 20 mg/kg). If the antibody is to act in the brain, a dosage of 50 mg/kg to 100 mg/kg is usually appropriate. Generally, partially human antibodies and fully human antibodies have a longer half-life within the human body than other antibodies. Accordingly, lower dosages and less frequent administration is often possible. Modifications such as lipidation can be used to stabilize antibodies and to enhance uptake and tissue penetration (e.g., into the brain). A method for lipidation of antibodies is described by Cruikshank et al. ((1997) J. Acquired Immune Deficiency Syndromes and Human Retrovirology 14:193).

The nucleic acid molecules of the invention can be inserted into vectors and used as gene therapy vectors. Gene therapy vectors can be delivered to a subject by, for example, intravenous injection, local administration (U.S. Patent 5,328,470), or by stereotactic injection (see, e.g., Chen et al. (1994) Proc. Natl. Acad. Sci. USA 91:3054-3057). The pharmaceutical preparation of the gene therapy vector can include the gene therapy vector in an acceptable diluent, or can comprise a slow release matrix in which the gene delivery vehicle is imbedded.

Alternatively, where the complete gene delivery vector can be produced intact from recombinant cells, e.g. retroviral vectors, the pharmaceutical preparation can include one or more cells which produce the gene delivery system.

The pharmaceutical compositions can be included in a container, pack, or dispenser together with instructions for administration.

25

30

5

10

15

20

V. Uses and Methods of the Invention

The nucleic acid molecules, proteins, protein homologs, and antibodies described herein can be used in one or more of the following methods: a) screening assays; b) detection assays (e.g., chromosomal mapping, tissue typing, forensic biology); c) predictive medicine (e.g., diagnostic assays, prognostic assays, monitoring clinical trials, and pharmacogenomics); and d) methods of treatment (e.g., therapeutic and prophylactic). For example, polypeptides of the invention can

to used for all of the purposes identified herein in portions of the disclosure relating to individual types of protein of the invention (e.g., TANGO 202 proteins, TANGO 234 proteins, TANGO 265 proteins, TANGO 273 proteins, TANGO 286 proteins, TANGO 294 proteins, and INTERCEPT 296 proteins). Polypeptides of the invention can also be used to modulate cellular proliferation, cellular differentiation, cellular adhesion, or some combination of these. The isolated nucleic acid molecules of the invention can be used to express proteins (e.g., via a recombinant expression vector in a host cell in gene therapy applications), to detect mRNA (e.g., in a biological sample) or a genetic lesion, and to modulate activity of a polypeptide of the invention. In addition, the polypeptides of the invention can be used to screen drugs or compounds which modulate activity or expression of a polypeptide of the invention as well as to treat disorders characterized by insufficient or excessive production of a protein of the invention or production of a form of a protein of the invention which has decreased or aberrant activity compared to the wild type protein. In addition, the antibodies of the invention can be used to detect and isolate a protein of the and modulate activity of a protein of the invention.

This invention further pertains to novel agents identified by the above-described screening assays and uses thereof for treatments as described herein.

20

25

30

5

10

15

A. Screening Assays

The invention provides a method (also referred to herein as a "screening assay") for identifying modulators, i.e., candidate or test compounds or agents (e.g., peptides, peptidomimetics, small molecules or other drugs) which bind to polypeptide of the invention or have a stimulatory or inhibitory effect on, for example, expression or activity of a polypeptide of the invention.

In one embodiment, the invention provides assays for screening candidate or test compounds which bind to or modulate the activity of the membrane-bound form of a polypeptide of the invention or biologically active portion thereof. The test compounds of the present invention can be obtained using any of the numerous approaches in combinatorial library methods known in the art, including: biological libraries; spatially addressable parallel solid phase or solution

PCT/US00/14858 WO 00/77239

phase libraries; synthetic library methods requiring deconvolution; the "one-bead one-compound" library method; and synthetic library methods using affinity chromatography selection. The biological library approach is limited to peptide libraries, while the other four approaches are applicable to peptide, non-peptide oligomer or small molecule libraries of compounds (Lam (1997) Anticancer Drug Des. 12:145).

5

10

15

25

30

Examples of methods for the synthesis of molecular libraries can be found in the art, for example in: DeWitt et al. (1993) Proc. Natl. Acad. Sci. USA 90:6909; Erb et al. (1994) Proc. Natl. Acad. Sci. USA 91:11422; Zuckermann et al. (1994). J. Med. Chem. 37:2678; Cho et al. (1993) Science 261:1303; Carrell et al. (1994) Angew. Chem. Int. Ed. Engl. 33:2059; Carell et al. (1994) Angew. Chem. Int. Ed. Engl. 33:2061; and Gallop et al. (1994) J. Med. Chem. 37:1233.

Libraries of compounds can be presented in solution (e.g., Houghten (1992) Bio/Techniques 13:412-421), or on beads (Lam (1991) Nature 354:82-84), chips (Fodor (1993) Nature 364:555-556), bacteria (U.S. Patent No. 5,223,409), spores (Patent Nos. 5,571,698; 5,403,484; and 5,223,409), plasmids (Cull et al. (1992) Proc. Natl. Acad. Sci. USA 89:1865-1869) or phage (Scott and Smith (1990) Science 249:386-390; Devlin (1990) Science 249:404-406; Cwirla et al. (1990) Proc. Natl. Acad. Sci. USA 87:6378-6382; and Felici (1991) J. Mol. Biol. 222:301-310). 20

In one embodiment, an assay is a cell-based assay in which a cell which expresses a membrane-bound form of a polypeptide of the invention, or a biologically active portion thereof, on the cell surface is contacted with a test compound and the ability of the test compound to bind to the polypeptide determined. The cell, for example, can be a yeast cell or a cell of mammalian origin. Determining the ability of the test compound to bind to the polypeptide can be accomplished, for example, by coupling the test compound with a radioisotope or enzymatic label such that binding of the test compound to the polypeptide or biologically active portion thereof can be determined by detecting the labeled compound in a complex. For example, test compounds can be labeled with 125I, 35S, ¹⁴C, or ³H, either directly or indirectly, and the radioisotope detected by direct counting of radio-emission or by scintillation counting. Alternatively, test

PCT/US00/14858 WO 00/77239

compounds can be enzymatically labeled with, for example, horseradish peroxidase, alkaline phosphatase, or luciferase, and the enzymatic label detected by determination of conversion of an appropriate substrate to product. In a preferred embodiment, the assay comprises contacting a cell which expresses a membranebound form of a polypeptide of the invention, or a biologically active portion thereof, on the cell surface with a known compound which binds the polypeptide to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with the polypeptide, wherein determining the ability of the test compound to interact with the polypeptide comprises determining the ability of the test compound to preferentially 10 bind to the polypeptide or a biologically active portion thereof as compared to the known compound.

5

15

20

25

30

In another embodiment, the assay involves assessment of an activity characteristic of the polypeptide, wherein binding of the test compound with the polypeptide or a biologically active portion thereof alters (i.e., increases or decreases) the activity of the polypeptide.

In another embodiment, an assay is a cell-based assay comprising contacting a cell expressing a membrane-bound form of a polypeptide of the invention, or a biologically active portion thereof, on the cell surface with a test compound and determining the ability of the test compound to modulate (e.g., stimulate or inhibit) the activity of the polypeptide or biologically active portion thereof. Determining the ability of the test compound to modulate the activity of the polypeptide or a biologically active portion thereof can be accomplished, for example, by determining the ability of the polypeptide to bind to or interact with a target molecule or to transport molecules across the cytoplasmic membrane.

Determining the ability of a polypeptide of the invention to bind to or interact with a target molecule can be accomplished by one of the methods described above for determining direct binding. As used herein, a "target molecule" is a molecule with which a selected polypeptide (e.g., a polypeptide of the invention binds or interacts with in nature, for example, a molecule on the surface of a cell which expresses the selected protein, a molecule on the surface of a second cell, a molecule in the extracellular milieu, a molecule associated with the internal surface

of a cell membrane or a cytoplasmic molecule. A target molecule can be a polypeptide of the invention or some other polypeptide or protein. For example, a target molecule can be a component of a signal transduction pathway which facilitates transduction of an extracellular signal (e.g., a signal generated by binding of a compound to a polypeptide of the invention) through the cell membrane and into the cell or a second intercellular protein which has catalytic activity or a protein which facilitates the association of downstream signaling molecules with a polypeptide of the invention. Determining the ability of a polypeptide of the invention to bind to or interact with a target molecule can be accomplished by determining the activity of the target molecule. For example, the activity of the target molecule can be determined by detecting induction of a cellular second messenger of the target (e.g., an mRNA, intracellular Ca2+, diacylglycerol, IP3, and the like), detecting catalytic/enzymatic activity of the target on an appropriate substrate, detecting the induction of a reporter gene (e.g., a regulatory element that is responsive to a polypeptide of the invention operably linked to a nucleic acid encoding a detectable marker, e.g. luciferase), or detecting a cellular response, for example, cellular differentiation, or cell proliferation.

5

10

15

20

25

30

In yet another embodiment, an assay of the present invention is a cell-free assay comprising contacting a polypeptide of the invention or biologically active portion thereof with a test compound and determining the ability of the test compound to bind to the polypeptide or biologically active portion thereof. Binding of the test compound to the polypeptide can be determined either directly or indirectly as described above. In a preferred embodiment, the assay includes contacting the polypeptide of the invention or biologically active portion thereof with a known compound which binds the polypeptide to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with the polypeptide, wherein determining the ability of the test compound to interact with the polypeptide comprises determining the ability of the test compound to preferentially bind to the polypeptide or biologically active portion thereof as compared to the known compound.

In another embodiment, an assay is a cell-free assay comprising contacting a polypeptide of the invention or biologically active portion thereof with

a test compound and determining the ability of the test compound to modulate (e.g., stimulate or inhibit) the activity of the polypeptide or biologically active portion thereof. Determining the ability of the test compound to modulate the activity of the polypeptide can be accomplished, for example, by determining the ability of the polypeptide to bind to a target molecule by one of the methods described above for determining direct binding. In an alternative embodiment, determining the ability of the test compound to modulate the activity of the polypeptide can be accomplished by determining the ability of the polypeptide of the invention to further modulate the target molecule. For example, the catalytic activity, the enzymatic activity, or both, of the target molecule on an appropriate substrate can be determined as previously described.

5

10

15

20

25

30

In yet another embodiment, the cell-free assay comprises contacting a polypeptide of the invention or biologically active portion thereof with a known compound which binds the polypeptide to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with the polypeptide, wherein determining the ability of the test compound to interact with the polypeptide comprises determining the ability of the polypeptide to preferentially bind to or modulate the activity of a target molecule.

The cell-free assays of the present invention are amenable to use of both a soluble form or the membrane-bound form of a polypeptide of the invention. In the case of cell-free assays comprising the membrane-bound form of the polypeptide, it can be desirable to utilize a solubilizing agent such that the membrane-bound form of the polypeptide is maintained in solution. Examples of such solubilizing agents include non-ionic detergents such as n-octylglucoside, n-dodecylglucoside, n-octylmaltoside, octanoyl-N-methylglucamide, decanoyl-N-methylglucamide, Triton X-100, Triton X-114, Thesit, Isotridecypoly(ethylene glycol ether)n, 3-[(3-cholamidopropyl)dimethylamminio]-1-propane sulfonate (CHAPSO), or N-dodecyl-N,N-dimethyl-3-ammonio-1-propane sulfonate.

In one or more embodiments of the above assay methods of the present invention, it can be desirable to immobilize either the polypeptide of the

5

10

15

20

25

30

invention or its target molecule to facilitate separation of complexed from noncomplexed forms of one or both of the proteins, as well as to accommodate automation of the assay. Binding of a test compound to the polypeptide, or interaction of the polypeptide with a target molecule in the presence and absence of a candidate compound, can be accomplished in any vessel suitable for containing the reactants. Examples of such vessels include microtiter plates, test tubes, and micro-centrifuge tubes. In one embodiment, a fusion protein can be provided which adds a domain that allows one or both of the proteins to be bound to a matrix. For example, glutathione-S-transferase fusion proteins or glutathione-S-transferase fusion proteins can be adsorbed onto glutathione SEPHAROSE™ beads (Sigma Chemical; St. Louis, MO) or glutathione derivatized microtiter plates, which are then combined with the test compound or the test compound and either the nonadsorbed target protein or A polypeptide of the invention, and the mixture incubated under conditions conducive to complex formation (e.g., at physiological conditions for salt and pH). Following incubation, the beads or microtiter plate wells are washed to remove any unbound components and complex formation is measured either directly or indirectly, for example, as described above. Alternatively, the complexes can be dissociated from the matrix, and the level of binding or activity of the polypeptide of the invention can be determined using standard techniques.

Other techniques for immobilizing proteins on matrices can also be used in the screening assays of the invention. For example, either the polypeptide of the invention or its target molecule can be immobilized utilizing conjugation of biotin and streptavidin. Biotinylated polypeptide of the invention or target molecules can be prepared from biotin-NHS (N-hydroxy-succinimide) using techniques well known in the art (e.g., biotinylation kit, Pierce Chemicals; Rockford, IL), and immobilized in the wells of streptavidin-coated 96 well plates (Pierce Chemical). Alternatively, antibodies reactive with the polypeptide of the invention or target molecules but which do not interfere with binding of the polypeptide of the invention to its target molecule can be derivatized to the wells of the plate, and unbound target or polypeptide of the invention trapped in the wells by antibody conjugation. Methods for detecting such complexes, in addition to those described above for the GST-immobilized complexes, include immunodetection of

complexes using antibodies reactive with the polypeptide of the invention or target molecule, as well as enzyme-linked assays which rely on detecting an enzymatic activity associated with the polypeptide of the invention or target molecule.

5

10

15

20

25

30

In another embodiment, modulators of expression of a polypeptide of the invention are identified in a method in which a cell is contacted with a candidate compound and the expression of the selected mRNA or protein (i.e., the mRNA or protein corresponding to a polypeptide or nucleic acid of the invention) in the cell is determined. The level of expression of the selected mRNA or protein in the presence of the candidate compound is compared to the level of expression of the selected mRNA or protein in the absence of the candidate compound. The candidate compound can then be identified as a modulator of expression of the polypeptide of the invention based on this comparison. For example, when expression of the selected mRNA or protein is greater (i.e., statistically significantly greater) in the presence of the candidate compound than in its absence, the candidate compound is identified as a stimulator of the selected mRNA or protein expression. Alternatively, when expression of the selected mRNA or protein is less (i.e., statistically significantly less) in the presence of the candidate compound than in its absence, the candidate compound is identified as an inhibitor of the selected mRNA or protein expression. The level of the selected mRNA or protein expression in the cells can be determined by methods described herein.

In yet another aspect of the invention, a polypeptide of the inventions can be used as "bait proteins" in a two-hybrid assay or three hybrid assay (see, e.g., U.S. Patent No. 5,283,317; Zervos et al. (1993) Cell 72:223-232; Madura et al. (1993) J. Biol. Chem. 268:12046-12054; Bartel et al. (1993) Bio/Techniques 14:920-924; Iwabuchi et al. (1993) Oncogene 8:1693-1696; and PCT Publication No. WO 94/10300), to identify other proteins, which bind to or interact with the polypeptide of the invention and modulate activity of the polypeptide of the invention. Such binding proteins are also likely to be involved in the propagation of signals by the polypeptide of the inventions as, for example, upstream or downstream elements of a signaling pathway involving the polypeptide of the invention.

This invention further pertains to novel agents identified by the above-described screening assays and uses thereof for treatments as described herein.

5 B. Detection Assays

10

15

20

25

30

Portions or fragments of the cDNA sequences identified herein (and the corresponding complete gene sequences) can be used in numerous ways as polynucleotide reagents. For example, these sequences can be used to: (i) map their respective genes on a chromosome and, thus, locate gene regions associated with genetic disease; (ii) identify an individual from a minute biological sample (tissue typing); and (iii) aid in forensic identification of a biological sample. These applications are described in the subsections below.

1. Chromosome Mapping

Once the sequence (or a portion of the sequence) of a gene has been isolated, this sequence can be used to map the location of the gene on a chromosome. Accordingly, nucleic acid molecules described herein or fragments thereof, can be used to map the location of the corresponding genes on a chromosome. The mapping of the sequences to chromosomes is an important first step in correlating these sequences with genes associated with disease.

Briefly, genes can be mapped to chromosomes by preparing PCR primers (preferably 15-25 base pairs in length) from the sequence of a gene of the invention. Computer analysis of the sequence of a gene of the invention can be used to rapidly select primers that do not span more than one exon in the genomic DNA, thus complicating the amplification process. These primers can then be used for PCR screening of somatic cell hybrids containing individual human chromosomes. Only those hybrids containing the human gene corresponding to the gene sequences will yield an amplified fragment. For a review of this technique, see D'Eustachio et al. ((1983) Science 220:919-924).

PCR mapping of somatic cell hybrids is a rapid procedure for assigning a particular sequence to a particular chromosome. Three or more sequences can be assigned per day using a single thermal cycler. Using the nucleic

acid sequences of the invention to design oligonucleotide primers, sub-localization can be achieved with panels of fragments from specific chromosomes. Other mapping strategies which can similarly be used to map a gene to its chromosome include *in situ* hybridization (described in Fan et al. (1990) *Proc. Natl. Acad. Sci. USA* 87:6223-27), pre-screening with labeled flow-sorted chromosomes, and pre-selection by hybridization to chromosome specific cDNA libraries. Fluorescence *in situ* hybridization (FISH) of a DNA sequence to a metaphase chromosomal spread can further be used to provide a precise chromosomal location in one step. For a review of this technique, see Verma et al. (Human Chromosomes: A Manual of Basic Techniques (Pergamon Press, New York, 1988)).

5

10

15

20

25

30

Reagents for chromosome mapping can be used individually to mark a single chromosome or a single site on that chromosome, or panels of reagents can be used for marking multiple sites and/or multiple chromosomes. Reagents corresponding to non-coding regions of the genes actually are preferred for mapping purposes. Coding sequences are more likely to be conserved within gene families, thus increasing the chance of cross hybridizations during chromosomal mapping.

Once a sequence has been mapped to a precise chromosomal location, the physical position of the sequence on the chromosome can be correlated with genetic map data. (Such data are found, for example, in V. McKusick, Mendelian Inheritance in Man, available on-line through Johns Hopkins University Welch Medical Library). The relationship between genes and disease, mapped to the same chromosomal region, can then be identified through linkage analysis (co-inheritance of physically adjacent genes), described in, e.g., Egeland et al. (1987) Nature 325:783-787.

Moreover, differences in the DNA sequences between individuals affected and unaffected with a disease associated with a gene of the invention can be determined. If a mutation is observed in some or all of the affected individuals but not in any unaffected individuals, then the mutation is likely to be the causative agent of the particular disease. Comparison of affected and unaffected individuals generally involves first looking for structural alterations in the chromosomes such as deletions or translocations that are visible from chromosome spreads or detectable using PCR based on that DNA sequence. Ultimately, complete

sequencing of genes from several individuals can be performed to confirm the presence of a mutation and to distinguish mutations from polymorphisms.

5

10

15

20

25

30

Furthermore, the nucleic acid sequences disclosed herein can be used to perform searches against "mapping databases", e.g., BLAST-type search, such that the chromosome position of the gene is identified by sequence homology or identity with known sequence fragments which have been mapped to chromosomes.

In the instant case, the human gene for TANGO 265 is located on chromosome 1 between markers D1S305 and D1S2635, and the human gene for TANGO 273 is located on chromosome 7 between markers D7S2467 and D7S2552.

In the instant case, the human gene for TANGO 286 exhibits significant amino acid homology with a region of the human chromosome region 22q12-13 genomic nucleotide sequence having GenBank Accession number AL021937. Alignment of a 45 kilobase nucleotide sequence encoding TANGO 286 with AL021937, however, indicated the presence in TANGO 286 of exons which differ from those disclosed in L021937 (pam120.mat scoring matrix; gap penalties -12/-4). This region of chromosome 22 comprises an immunoglobulin lambda chain C (IGLC) pseudogene, the Ret finger protein-like 3 (RFPL3) and Ret finger proteinlike 3 antisense (RFPL3S) genes, a gene encoding a novel immunoglobulin lambda chain V family protein, a novel gene encoding a protein similar both to mouse RGDS protein (RALGDS, RALGEF, guanine nucleotide dissociation stimulator A) and to rabbit oncogene RSC, a novel gene encoding the human orthologue of worm F16A11.2 protein, a novel gene encoding a protein similar both to BPI and to rabbit liposaccharide-binding protein, and a 5'-portion of a novel gene. This region also comprises various ESTs, STSs, GSSs, genomic marker D22S1175, a ca repeat polymorphism and putative CpG islands.

A polypeptide and fragments and sequences thereof and antibodies which bind specifically with such polypeptides/fragments can be used to map the location of the gene encoding the polypeptide on a chromosome. This mapping can be performed by specifically detecting the presence of the polypeptide/fragments in members of a panel of somatic cell hybrids between cells obtained from a first species of animal from which the protein originates and cells obtained from a

second species of animal, determining which somatic cell hybrid(s) expresses the polypeptide, and noting the chromosome(s) of the first species of animal that it contains. For examples of this technique (see Pajunen et al., 1988, Cytogenet. Cell Genet. 47:37-41 and Van Keuren et al., 1986, Hum. Genet. 74:34-40).

Alternatively, the presence of the polypeptide in the somatic cell hybrids can be determined by assaying an activity or property of the polypeptide (e.g., enzymatic activity, as described in Bordelon-Riser et al., 1979, Som. Cell Genet. 5:597-613 and Owerbach et al., 1978, Proc. Natl. Acad. Sci. USA 75:5640-5644).

In the instant case, the human gene for TANGO 234 protein
indicated that the gene is located at chromosomal location h12p13. Flanking
chromosomal markers include WI-6980 and GATA8A09.43. Nearby human loci
include IBD2 (inflammatory bowel disease 2), FPF (familial periodic fever), and
HPDR2 (hypophosphatemia vitamin D resistant rickets 2). Nearby genes are
KLRC (killer cell receptor cluster), DRPLA (dentatorubro-pallidoluysian atrophy),
GAPD (glyceraldehyde-3-phosphate) dehydrogenase, and PXR1 (peroxisome
receptor 1). This region is syntenic to mouse chromosome mo6. Murine
chromosomal mapping indicated that the murine orthologue is located near the scr
(scruffy) locus. Nearby mouse genes include drpla (dentatorubral phillidoluysian
atrophy), prp (proline rich protein), and kap (kidney androgen regulated protein).

20

25

30

2. Tissue Typing

The nucleic acid sequences of the present invention can also be used to identify individuals from minute biological samples. The United States military, for example, is considering the use of restriction fragment length polymorphism (RFLP) for identification of its personnel. In this technique, an individual's genomic DNA is digested with one or more restriction enzymes, and probed on a Southern blot to yield unique bands for identification. This method does not suffer from the current limitations of "Dog Tags" which can be lost, switched, or stolen, making positive identification difficult. The sequences of the present invention are useful as additional DNA markers for RFLP (described in U.S. Patent 5,272,057).

Furthermore, the sequences of the present invention can be used to provide an alternative technique which determines the actual base-by-base DNA

sequence of selected portions of an individual's genome. Thus, the nucleic acid sequences described herein can be used to prepare two PCR primers from the 5' and 3' ends of the sequences. These primers can then be used to amplify an individual's DNA and subsequently sequence it.

5

10

15

20

25

30

Panels of corresponding DNA sequences from individuals, prepared in this manner, can provide unique individual identifications, as each individual will have a unique set of such DNA sequences due to allelic differences. The sequences of the present invention can be used to obtain such identification sequences from individuals and from tissue. The nucleic acid sequences of the invention uniquely represent portions of the human genome. Allelic variation occurs to some degree in the coding regions of these sequences, and to a greater degree in the non-coding regions. It is estimated that allelic variation between individual humans occurs with a frequency of about once per each 500 bases. Each of the sequences described herein can, to some degree, be used as a standard against which DNA from an individual can be compared for identification purposes. Because greater numbers of polymorphisms occur in the non-coding regions, fewer sequences are necessary to differentiate individuals. The non-coding sequences of any of SEQ ID NOs: 1, 9, 17, 25, 33, 45, and 53 can comfortably provide positive individual identification with a panel of perhaps 10 to 1,000 primers which each yield a non-coding amplified sequence of 100 bases. If predicted coding sequences, such as those in any of SEQ ID NOs: 2, 10, 18, 26, 34, 46, and 54 are used, a more appropriate number of primers for positive individual identification would be 500-2,000.

If a panel of reagents from the nucleic acid sequences described herein is used to generate a unique identification database for an individual, those same reagents can later be used to identify tissue from that individual. Using the unique identification database, positive identification of the individual, living or dead, can be made from extremely small tissue samples.

3. Use of Partial Gene Sequences in Forensic Biology

DNA-based identification techniques can also be used in forensic biology. Forensic biology is a scientific field employing genetic typing of biological evidence found at a crime scene as a means for positively identifying, for

example, a perpetrator of a crime. To make such an identification, PCR technology can be used to amplify DNA sequences taken from very small biological samples such as tissues, e.g., hair or skin, or body fluids, e.g., blood, saliva, or semen found at a crime scene. The amplified sequence can then be compared to a standard, thereby allowing identification of the origin of the biological sample.

The sequences of the present invention can be used to provide polynucleotide reagents, e.g., PCR primers, targeted to specific loci in the human genome, which can enhance the reliability of DNA-based forensic identifications by, for example, providing another "identification marker" (i.e., another DNA sequence that is unique to a particular individual). As mentioned above, actual base sequence information can be used for identification as an accurate alternative to patterns formed by restriction enzyme generated fragments. Sequences targeted to non-coding regions are particularly appropriate for this use as greater numbers of polymorphisms occur in the non-coding regions, making it easier to differentiate individuals using this technique. Examples of polynucleotide reagents include the nucleic acid sequences of the invention or portions thereof, e.g., fragments derived from non-coding regions having a length of at least 20 or 30 bases.

The nucleic acid sequences described herein can further be used to provide polynucleotide reagents, e.g., labeled or labelable probes which can be used in, for example, an *in situ* hybridization technique, to identify a specific tissue, e.g., brain tissue. This can be very useful in cases where a forensic pathologist is presented with a tissue of unknown origin. Panels of such probes can be used to identify tissue by species and/or by organ type.

25 C. Predictive Medicine

5

10

15

20

30

The present invention also pertains to the field of predictive medicine in which diagnostic assays, prognostic assays, pharmacogenomics, and monitoring clinical trials are used for prognostic (predictive) purposes to thereby treat an individual prophylactically. Accordingly, one aspect of the present invention relates to diagnostic assays for determining expression of a polypeptide or nucleic acid of the invention and/or activity of a polypeptide of the invention (e.g., expression or activity of one of TANGO 202, TANGO 234, TANGO 265, TANGO

273, TANGO 286, TANGO 294, or INTERCEPT 296 genes or proteins), in the context of a biological sample (e.g., blood, serum, cells, tissue) to thereby determine whether an individual is afflicted with a disease or disorder, or is at risk of developing a disorder, associated with aberrant expression or activity of a polypeptide of the invention. The invention also provides for prognostic (or predictive) assays for determining whether an individual is at risk of developing a disorder associated with aberrant expression or activity of a polypeptide of the invention. For example, mutations in a gene of the invention can be assayed in a biological sample. Such assays can be used for prognostic or predictive purpose to thereby prophylactically treat an individual prior to the onset of a disorder characterized by or associated with aberrant expression or activity of a polypeptide of the invention.

5

10

15

20

25

30

As an alternative to making determinations based on the absolute expression level of a selected gene, determinations can be based on normalized expression levels of the gene. A gene expression level is normalized by correcting the absolute expression level of the gene (e.g., a TANGO 202, TANGO 234, TANGO 265, TANGO 273, TANGO 286, TANGO 294, or INTERCEPT 296 gene as described herein) by comparing its expression to expression of a gene for which expression is not believed to be co-regulated with the gene of interest, e.g., a housekeeping gene that is constitutively expressed. Suitable genes for normalization include housekeeping genes such as the actin gene. Such normalization allows comparison of the expression level in one sample, e.g., a patient sample, with the expression level in another sample, e.g., a sample obtained from a patient known not to be afflicted with a disease or condition, or between samples obtained from different sources.

Alternatively, the expression level can be assessed as a relative expression level. To assess a relative expression level for a gene (e.g., a TANGO 202, TANGO 234, TANGO 265, TANGO 273, TANGO 286, TANGO 294, or INTERCEPT 296 gene, as described herein), the level of expression of the gene is determined for 10 or more samples (preferably 50 or more samples) of different isolates of cells in which the gene is believed to be expressed, prior to assessing the level of expression of the gene in the sample of interest. The mean expression level

of the gene detected in the large number of samples is determined, and this value is used as a baseline expression level for the gene. The expression level of the gene assessed in the test sample (i.e., its absolute level of expression) is divided by the mean expression value to yield a relative expression level. Such a method can identify tissues or individuals which are afflicted with a disorder associated with aberrant expression of a gene of the invention.

Preferably, the samples used in the baseline determination are generated either using cells obtained from a tissue or individual known to be afflicted with a disorder (e.g., a disorder associated with aberrant expression of one of the TANGO 202, TANGO 234, TANGO 265, TANGO 273, TANGO 286, TANGO 294, or INTERCEPT 296 genes) or using cells obtained from a tissue or individual known not to be afflicted with the disorder. Alternatively, levels of expression of these genes in tissues or individuals known to be or not to be afflicted with the disorder can be used to assess whether the aberrant expression of the gene is associated with the disorder (e.g., with onset of the disorder, or as a symptom of the disorder over time).

Another aspect of the invention pertains to monitoring the influence of agents (e.g., drugs or other compounds) on the expression or activity of one or more of TANGO 202, TANGO 234, TANGO 265, TANGO 273, TANGO 286, TANGO 294, and INTERCEPT 296 in clinical trials. These and other agents are described in further detail in the following sections.

1. Diagnostic Assays

5

10

15

20

25

30

An exemplary method for detecting the presence or absence of a polypeptide or nucleic acid of the invention in a biological sample involves obtaining a biological sample from a test subject and contacting the biological sample with a compound or an agent capable of detecting a polypeptide or nucleic acid (e.g., mRNA, genomic DNA) of the invention such that the presence of a polypeptide or nucleic acid of the invention is detected in the biological sample. A preferred agent for detecting mRNA or genomic DNA encoding a polypeptide of the invention is a labeled nucleic acid probe capable of hybridizing to mRNA or genomic DNA encoding a polypeptide of the invention. The nucleic acid probe can

be, for example, a full-length cDNA, such as the nucleic acid of any of SEQ ID NOs: 1, 9, 17, 25, 33, 45, 53, 67, and 72, or a portion thereof, such as an oligonucleotide of at least 15, 30, 50, 100, 250 or 500 nucleotides in length and sufficient to specifically hybridize under stringent conditions to a mRNA or genomic DNA encoding a polypeptide of the invention. Other suitable probes for use in the diagnostic assays of the invention are described herein.

5

A preferred agent for detecting a polypeptide of the invention is an antibody capable of binding to a polypeptide of the invention, preferably an antibody with a detectable label. Antibodies can be polyclonal, or more preferably, monoclonal. An intact antibody, or a fragment thereof (e.g., Fab or F(ab')2) can be 10 used. The term "labeled", with regard to the probe or antibody, is intended to encompass direct labeling of the probe or antibody by coupling (i.e., physically linking) a detectable substance to the probe or antibody, as well as indirect labeling of the probe or antibody by reactivity with another reagent that is directly labeled. Examples of indirect labeling include detection of a primary antibody using a 15 fluorescently labeled secondary antibody and end-labeling of a DNA probe with biotin such that it can be detected with fluorescently labeled streptavidin. The term "biological sample" is intended to include tissues, cells and biological fluids isolated from a subject, as well as tissues, cells and fluids present within a subject. That is, the detection method of the invention can be used to detect mRNA, protein, 20 or genomic DNA in a biological sample in vitro as well as in vivo. For example, in vitro techniques for detection of mRNA include Northern hybridizations and in situ hybridizations. In vitro techniques for detection of a polypeptide of the invention include enzyme linked immunosorbent assays (ELISAs), Western blots, immunoprecipitations and immunofluorescence. In vitro techniques for detection 25 of genomic DNA include Southern hybridizations. Furthermore, in vivo techniques for detection of a polypeptide of the invention include introducing into a subject a labeled antibody directed against the polypeptide. For example, the antibody can be labeled with a radioactive marker whose presence and location in a subject can be detected by standard imaging techniques. 30

In one embodiment, the biological sample contains protein molecules from the test subject. Alternatively, the biological sample can contain

PCT/US00/14858 WO 00/77239

mRNA molecules from the test subject or genomic DNA molecules from the test subject. A preferred biological sample is a peripheral blood leukocyte sample isolated by conventional means from a subject.

5

10

15

25

30

In another embodiment, the methods further involve obtaining a control biological sample from a control subject, contacting the control sample with a compound or agent capable of detecting a polypeptide of the invention or mRNA or genomic DNA encoding a polypeptide of the invention, such that the presence of the polypeptide or mRNA or genomic DNA encoding the polypeptide is detected in the biological sample, and comparing the presence of the polypeptide or mRNA or genomic DNA encoding the polypeptide in the control sample with the presence of the polypeptide or mRNA or genomic DNA encoding the polypeptide in the test sample.

The invention also encompasses kits for detecting the presence of a polypeptide or nucleic acid of the invention in a biological sample (a test sample). Such kits can be used to determine if a subject is suffering from or is at increased risk of developing a disorder associated with aberrant expression of a polypeptide of the invention (e.g., one of the disorders described in the section of this disclosure wherein the individual polypeptide of the invention is discussed). For example, the kit can comprise a labeled compound or agent capable of detecting the polypeptide 20 or mRNA encoding the polypeptide in a biological sample and means for determining the amount of the polypeptide or mRNA in the sample (e.g., an antibody which binds the polypeptide or an oligonucleotide probe which binds to DNA or mRNA encoding the polypeptide). Kits can also include instructions for observing that the tested subject is suffering from or is at risk of developing a disorder associated with aberrant expression of the polypeptide if the amount of the polypeptide or mRNA encoding the polypeptide is above or below a normal level.

For antibody-based kits, the kit can comprise, for example: (1) a first antibody (e.g., attached to a solid support) which binds to a polypeptide of the invention; and, optionally, (2) a second, different antibody which binds to either the polypeptide or the first antibody and is conjugated to a detectable agent.

For oligonucleotide-based kits, the kit can comprise, for example: (1) an oligonucleotide, e.g., a detectably labeled oligonucleotide, which hybridizes to a

nucleic acid sequence encoding a polypeptide of the invention or (2) a pair of primers useful for amplifying a nucleic acid molecule encoding a polypeptide of the invention. The kit can also comprise, e.g., a buffering agent, a preservative, or a protein stabilizing agent. The kit can also comprise components necessary for detecting the detectable agent (e.g., an enzyme or a substrate). The kit can also contain a control sample or a series of control samples which can be assayed and compared to the test sample contained. Each component of the kit is usually enclosed within an individual container and all of the various containers are within a single package along with instructions for observing whether the tested subject is suffering from or is at risk of developing a disorder associated with aberrant expression of the polypeptide.

2. Prognostic Assays

5

10

15

20

25

30

The methods described herein can furthermore be utilized as diagnostic or prognostic assays to identify subjects having or at risk of developing a disease or disorder associated with aberrant expression or activity of a polypeptide of the invention. For example, the assays described herein, such as the preceding diagnostic assays or the following assays, can be utilized to identify a subject having or at risk of developing a disorder associated with aberrant expression or activity of a polypeptide of the invention (e.g., one of the disorders described in the section of this disclosure wherein the individual polypeptide of the invention is discussed). Alternatively, the prognostic assays can be utilized to identify a subject having or at risk for developing such a disease or disorder. Thus, the present invention provides a method in which a test sample is obtained from a subject and a polypeptide or nucleic acid (e.g., mRNA, genomic DNA) of the invention is detected, wherein the presence of the polypeptide or nucleic acid is diagnostic for a subject having or at risk of developing a disease or disorder associated with aberrant expression or activity of the polypeptide. As used herein, a "test sample" refers to a biological sample obtained from a subject of interest. For example, a test sample can be a biological fluid (e.g., serum), cell sample, or tissue.

Furthermore, the prognostic assays described herein can be used to determine whether a subject can be administered an agent (e.g., an agonist,

antagonist, peptidomimetic, protein, peptide, nucleic acid, small molecule, or other drug candidate) to treat a disease or disorder associated with aberrant expression or activity of a polypeptide of the invention. For example, such methods can be used to determine whether a subject can be effectively treated with a specific agent or class of agents (e.g., agents of a type which decrease activity of the polypeptide). Thus, the present invention provides methods for determining whether a subject can be effectively treated with an agent for a disorder associated with aberrant expression or activity of a polypeptide of the invention in which a test sample is obtained and the polypeptide or nucleic acid encoding the polypeptide is detected (e.g., wherein the presence of the polypeptide or nucleic acid is diagnostic for a subject that can be administered the agent to treat a disorder associated with aberrant expression or activity of the polypeptide).

10

15

20

25

30

The methods of the invention can also be used to detect genetic lesions or mutations in a gene of the invention, thereby determining if a subject with the lesioned gene is at risk for a disorder characterized aberrant expression or activity of a polypeptide of the invention. In preferred embodiments, the methods include detecting, in a sample of cells from the subject, the presence or absence of a genetic lesion or mutation characterized by at least one of an alteration affecting the integrity of a gene encoding the polypeptide of the invention, or the mis-expression of the gene encoding the polypeptide of the invention. For example, such genetic lesions or mutations can be detected by ascertaining the existence of at least one of: 1) a deletion of one or more nucleotides from the gene; 2) an addition of one or more nucleotides to the gene; 3) a substitution of one or more nucleotides of the gene; 4) a chromosomal rearrangement of the gene; 5) an alteration in the level of a messenger RNA transcript of the gene; 6) an aberrant modification of the gene, such as of the methylation pattern of the genomic DNA; 7) the presence of a non-wild type splicing pattern of a messenger RNA transcript of the gene; 8) a non-wild type level of the protein encoded by the gene; 9) an allelic loss of the gene; and 10) an inappropriate post-translational modification of the protein encoded by the gene. As described herein, there are a large number of assay techniques known in the art which can be used for detecting lesions in a gene.

In certain embodiments, detection of the lesion involves the use of a probe/primer in a polymerase chain reaction (PCR) (see, e.g., U.S. Patent Nos. 4,683,195 and 4,683,202), such as anchor PCR or RACE PCR, or, alternatively, in a ligation chain reaction (LCR) (see, e.g., Landegran et al. (1988) Science 241:1077-1080; and Nakazawa et al. (1994) Proc. Natl. Acad. Sci. USA 91:360-364), the latter of which can be particularly useful for detecting point mutations in a gene (see, e.g., Abravaya et al. (1995) Nucleic Acids Res. 23:675-682). This method can include the steps of collecting a sample of cells from a patient, isolating nucleic acid (e.g., genomic, mRNA or both) from the cells of the sample, contacting the nucleic acid sample with one or more primers which specifically hybridize to the selected gene under conditions such that hybridization and amplification of the gene (if present) occurs, and detecting the presence or absence of an amplification product, or detecting the size of the amplification product and comparing the length to a control sample. PCR and/or LCR can be desirable to use as a preliminary amplification step in conjunction with any of the techniques used for detecting mutations described herein.

5

10

15

20

25

30

Alternative amplification methods include: self-sustained sequence replication (Guatelli et al. (1990) *Proc. Natl. Acad. Sci. USA* 87:1874-1878), transcriptional amplification system (Kwoh, et al. (1989) *Proc. Natl. Acad. Sci. USA* 86:1173-1177), Q-Beta Replicase (Lizardi et al. (1988) *Bio/Technology* 6:1197), or any other nucleic acid amplification method, followed by the detection of the amplified molecules using techniques well known to those of skill in the art. These detection schemes are especially useful for the detection of nucleic acid molecules if such molecules are present in very low numbers.

In an alternative embodiment, mutations in a selected gene from a sample cell can be identified by alterations in restriction enzyme cleavage patterns. For example, sample and control DNA is isolated, (optionally) amplified, digested with one or more restriction endonucleases, and fragment length sizes are determined by gel electrophoresis and compared. Differences in fragment length sizes between sample and control DNA indicates mutations in the sample DNA. Moreover, the use of sequence specific ribozymes (see, e.g., U.S. Patent No.

5,498,531) can be used to score for the presence of specific mutations by development or loss of a ribozyme cleavage site.

5

10

15

20

25

In other embodiments, genetic mutations can be identified by hybridizing a sample and control nucleic acids, e.g., DNA or RNA, to high density arrays containing hundreds or thousands of oligonucleotides probes (Cronin et al. (1996) *Human Mutation* 7:244-255; Kozal et al. (1996) *Nature Medicine* 2:753-759). For example, genetic mutations can be identified in two-dimensional arrays containing light-generated DNA probes as described in Cronin et al., *supra*. Briefly, a first hybridization array of probes can be used to scan through long stretches of DNA in a sample and control to identify base changes between the sequences by making linear arrays of sequential overlapping probes. This step allows the identification of point mutations. This step is followed by a second hybridization array that allows the characterization of specific mutations by using smaller, specialized probe arrays complementary to all variants or mutations detected. Each mutation array is composed of parallel probe sets, one complementary to the wild-type gene and the other complementary to the mutant gene.

In yet another embodiment, any of a variety of sequencing reactions known in the art can be used to directly sequence the selected gene and detect mutations by comparing the sequence of the sample nucleic acids with the corresponding wild-type (control) sequence. Examples of sequencing reactions include those based on techniques developed by Maxim and Gilbert ((1977) Proc. Natl. Acad. Sci. USA 74:560) or Sanger ((1977) Proc. Natl. Acad. Sci. USA 74:5463). It is also contemplated that any of a variety of automated sequencing procedures can be utilized when performing the diagnostic assays ((1995) Bio/Techniques 19:448), including sequencing by mass spectrometry (see, e.g., PCT Publication No. WO 94/16101; Cohen et al. (1996) Adv. Chromatogr. 36:127-162; and Griffin et al. (1993) Appl. Biochem. Biotechnol. 38:147-159).

Other methods for detecting mutations in a selected gene include methods in which protection from cleavage agents is used to detect mismatched bases in RNA/RNA or RNA/DNA heteroduplexes (Myers et al. (1985) Science 230:1242). In general, the technique of mismatch cleavage entails providing

heteroduplexes formed by hybridizing (labeled) RNA or DNA containing the wild-type sequence with potentially mutant RNA or DNA obtained from a tissue sample. The double-stranded duplexes are treated with an agent which cleaves single-stranded regions of the duplex such as which will exist due to base pair mismatches between the control and sample strands. RNA/DNA duplexes can be treated with RNase to digest mismatched regions, and DNA/DNA hybrids can be treated with S1 nuclease to digest mismatched regions.

5

10

15

20

25

30

In other embodiments, either DNA/DNA or RNA/DNA duplexes can be treated with hydroxylamine or osmium tetroxide and with piperidine in order to digest mismatched regions. After digestion of the mismatched regions, the resulting material is then separated by size on denaturing polyacrylamide gels to determine the site of mutation. See, e.g., Cotton et al. (1988) Proc. Natl. Acad. Sci. USA 85:4397; Saleeba et al. (1992) Methods Enzymol. 217:286-295. In a preferred embodiment, the control DNA or RNA can be labeled for detection.

In still another embodiment, the mismatch cleavage reaction employs one or more proteins that recognize mismatched base pairs in double-stranded DNA (so called DNA mismatch repair enzymes) in defined systems for detecting and mapping point mutations in cDNAs obtained from samples of cells. For example, the mutY enzyme of E. coli cleaves A at G/A mismatches and the thymidine DNA glycosylase from HeLa cells cleaves T at G/T mismatches (Hsu et al. (1994) Carcinogenesis 15:1657-1662). According to an exemplary embodiment, a probe based on a selected sequence, e.g., a wild-type sequence, is hybridized to a cDNA or other DNA product from a test cell(s). The duplex is treated with a DNA mismatch repair enzyme, and the cleavage products, if any, can be detected from electrophoresis protocols or the like. See, e.g., U.S. Patent No. 5,459,039.

In other embodiments, alterations in electrophoretic mobility will be used to identify mutations in genes. For example, single strand conformation polymorphism (SSCP) can be used to detect differences in electrophoretic mobility between mutant and wild type nucleic acids (Orita et al. (1989) *Proc. Natl. Acad. Sci. USA* 86:2766; see also Cotton (1993) *Mutat. Res.* 285:125-144; Hayashi (1992) *Genet. Anal. Tech. Appl.* 9:73-79). Single-stranded DNA fragments of sample and control nucleic acids will be denatured and allowed to re-nature. The secondary

structure of single-stranded nucleic acids varies according to sequence, and the resulting alteration in electrophoretic mobility enables the detection of even a single base change. The DNA fragments can be labeled or detected with labeled probes. The sensitivity of the assay can be enhanced by using RNA (rather than DNA), in which the secondary structure is more sensitive to a change in sequence. In a preferred embodiment, the subject method utilizes heteroduplex analysis to separate double stranded heteroduplex molecules on the basis of changes in electrophoretic mobility (Keen et al. (1991) *Trends Genet.* 7:5).

5

10

15

20

25

30

In yet another embodiment, the movement of mutant or wild-type fragments in polyacrylamide gels containing a gradient of denaturant is assayed using denaturing gradient gel electrophoresis (DGGE) (Myers et al. (1985) *Nature* 313:495). When DGGE is used as the method of analysis, DNA will be modified to insure that it does not completely denature, for example by adding a 'GC clamp' of approximately 40 base pairs of high-melting GC-rich DNA by PCR. In a further embodiment, a temperature gradient is used in place of a denaturing gradient to identify differences in the mobility of control and sample DNA (Rosenbaum and Reissner (1987) *Biophys. Chem.* 265:12753).

Examples of other techniques for detecting point mutations include, but are not limited to, selective oligonucleotide hybridization, selective amplification, or selective primer extension. For example, oligonucleotide primers can be prepared in which the known mutation is placed centrally and then hybridized to target DNA under conditions which permit hybridization only if a perfect match is found (Saiki et al. (1986) *Nature* 324:163); Saiki et al. (1989) *Proc. Natl. Acad. Sci. USA* 86:6230). Such allele specific oligonucleotides are hybridized to PCR amplified target DNA or a number of different mutations when the oligonucleotides are attached to the hybridizing membrane and hybridized with labeled target DNA.

Alternatively, allele specific amplification technology which depends on selective PCR amplification can be used in conjunction with the instant invention. Oligonucleotides used as primers for specific amplification can carry the mutation of interest in the center of the molecule (so that amplification depends on differential hybridization; Gibbs et al. (1989) *Nucleic Acids Res.* 17:2437-2448) or

at the extreme 3' end of one primer where, under appropriate conditions, mismatching can prevent or reduce polymerase extension (Prossner (1993) *Tibtech* 11:238). In addition, it can be desirable to introduce a novel restriction site in the region of the mutation to create cleavage-based detection (Gasparini et al. (1992) *Mol. Cell Probes* 6:1). Amplification can also be performed using Taq ligase for amplification (Barany (1991) *Proc. Natl. Acad. Sci. USA* 88:189). In such cases, ligation will occur only if there is a perfect match at the 3' end of the 5' sequence making it possible to detect the presence of a known mutation at a specific site by looking for the presence or absence of amplification.

The methods described herein can be performed, for example, using pre-packaged diagnostic kits comprising at least one probe nucleic acid or antibody reagent described herein, which can be conveniently used, e.g., in clinical settings to diagnose patients exhibiting symptoms or family history of a disease or illness involving a gene encoding a polypeptide of the invention. Furthermore, any cell type or tissue, preferably peripheral blood leukocytes, in which the polypeptide of the invention is expressed can be utilized in the prognostic assays described herein.

3. Pharmacogenomics

5

10

15

20

25

30

Agents, or modulators which have a stimulatory or inhibitory effect on activity or expression of a polypeptide of the invention as identified by a screening assay described herein can be administered to individuals to treat (prophylactically or therapeutically) disorders associated with aberrant activity of the polypeptide. In conjunction with such treatment, the pharmacogenomics (i.e., the study of the relationship between an individual's genotype and that individual's response to a foreign compound or drug) of the individual may be considered. Differences in metabolism of therapeutics can lead to severe toxicity or therapeutic failure by altering the relation between dose and blood concentration of the pharmacologically active drug. Thus, the pharmacogenomics of the individual permits the selection of effective agents (e.g., drugs) for prophylactic or therapeutic treatments based on a consideration of the individual's genotype. Such pharmacogenomics can further be used to determine appropriate dosages and therapeutic regimens. Accordingly, the activity of a polypeptide of the invention,

expression of a nucleic acid of the invention, or mutation content of a gene of the invention in an individual can be determined to thereby select appropriate agent(s) for therapeutic or prophylactic treatment of the individual.

5

10

15

20

25

30

Pharmacogenomics deals with clinically significant hereditary variations in the response to drugs due to altered drug disposition and abnormal action in affected persons. See, e.g., Linder (1997) Clin. Chem. 43(2):254-266. In general, two types of pharmacogenetic conditions can be differentiated. Genetic conditions transmitted as a single factor altering the way drugs act on the body are referred to as "altered drug action." Genetic conditions transmitted as single factors altering the way the body acts on drugs are referred to as "altered drug metabolism". These pharmacogenetic conditions can occur either as rare defects or as polymorphisms. For example, glucose-6-phosphate dehydrogenase (G6PD) deficiency is a common inherited enzymopathy in which the main clinical complication is hemolysis after ingestion of oxidant drugs (anti-malarials, sulfonamides, analgesics, nitrofurans) and consumption of fava beans.

As an illustrative embodiment, the activity of drug metabolizing enzymes is a major determinant of both the intensity and duration of drug action. The discovery of genetic polymorphisms of drug metabolizing enzymes (e.g., Nacetyltransferase 2 (NAT 2) and cytochrome P450 enzymes CYP2D6 and CYP2C19) has provided an explanation as to why some patients do not obtain the expected drug effects or show exaggerated drug response and serious toxicity after taking the standard and safe dose of a drug. These polymorphisms are expressed in two phenotypes in the population, the extensive metabolizer (EM) and poor metabolizer (PM). The prevalence of PM is different among different populations. For example, the gene coding for CYP2D6 is highly polymorphic and several mutations have been identified in PM, which all lead to the absence of functional CYP2D6. Poor metabolizers of CYP2D6 and CYP2C19 quite frequently experience exaggerated drug response and side effects when they receive standard doses. If a metabolite is the active therapeutic moiety, a PM will show no therapeutic response, as demonstrated for the analgesic effect of codeine mediated by its CYP2D6-formed metabolite morphine. The other extreme are the so called ultra-rapid metabolizers who do not respond to standard doses. Recently, the

molecular basis of ultra-rapid metabolism has been identified to be due to CYP2D6 gene amplification.

Thus, the activity of a polypeptide of the invention, expression of a nucleic acid encoding the polypeptide, or mutation content of a gene encoding the polypeptide in an individual can be determined to thereby select appropriate agent(s) for therapeutic or prophylactic treatment of the individual. In addition, pharmacogenetic studies can be used to apply genotyping of polymorphic alleles encoding drug-metabolizing enzymes to the identification of an individual's drug responsiveness phenotype. This knowledge, when applied to dosing or drug selection, can avoid adverse reactions or therapeutic failure and thus enhance therapeutic or prophylactic efficiency when treating a subject with a modulator of activity or expression of the polypeptide, such as a modulator identified by one of the exemplary screening assays described herein.

4. Monitoring of Effects During Clinical Trials

5

10

15

20

25

30

Monitoring the influence of agents (e.g., drug compounds) on the expression or activity of a polypeptide of the invention (e.g., the ability to modulate aberrant cell proliferation chemotaxis, and/or differentiation) can be applied not only in basic drug screening, but also in clinical trials. For example, the effectiveness of an agent, as determined by a screening assay as described herein, to increase gene expression, protein levels, or protein activity, can be monitored in clinical trials of subjects exhibiting decreased gene expression, protein levels, or protein activity. Alternatively, the effectiveness of an agent, as determined by a screening assay, to decrease gene expression, protein levels or protein activity, can be monitored in clinical trials of subjects exhibiting increased gene expression, protein levels, or protein activity. In such clinical trials, expression or activity of a polypeptide of the invention and preferably, that of other polypeptide that have been implicated in for example, a cellular proliferation disorder, can be used as a marker of the immune responsiveness of a particular cell.

For example, and not by way of limitation, genes, including those of the invention, that are modulated in cells by treatment with an agent (e.g., compound, drug or small molecule) which modulates activity or expression of a

polypeptide of the invention (e.g., as identified in a screening assay described herein) can be identified. Thus, to study the effect of agents on cellular proliferation disorders, for example, in a clinical trial, cells can be isolated and RNA prepared and analyzed for the levels of expression of a gene of the invention and other genes implicated in the disorder. The levels of gene expression (i.e., a gene expression pattern) can be quantified by Northern blot analysis or RT-PCR, as described herein, or alternatively by measuring the amount of protein produced, by one of the methods as described herein, or by measuring the levels of activity of a gene of the invention or other genes. In this way, the gene expression pattern can serve as a marker, indicative of the physiological response of the cells to the agent. Accordingly, this response state can be determined before, and at various points during, treatment of the individual with the agent.

5

10

15

20

25

30

In a preferred embodiment, the present invention provides a method for monitoring the effectiveness of treatment of a subject with an agent (e.g., an agonist, antagonist, peptidomimetic, protein, peptide, nucleic acid, small molecule, or other drug candidate identified by the screening assays described herein) comprising the steps of (i) obtaining a pre-administration sample from a subject prior to administration of the agent; (ii) detecting the level of the polypeptide or nucleic acid of the invention in the pre-administration sample; (iii) obtaining one or more post-administration samples from the subject; (iv) detecting the level the of the polypeptide or nucleic acid of the invention in the post-administration samples; (v) comparing the level of the polypeptide or nucleic acid of the invention in the pre-administration sample with the level of the polypeptide or nucleic acid of the invention in the post-administration sample or samples; and (vi) altering the administration of the agent to the subject accordingly. For example, increased administration of the agent can be desirable to increase the expression or activity of the polypeptide to higher levels than detected, i.e., to increase the effectiveness of the agent. Alternatively, decreased administration of the agent can be desirable to decrease expression or activity of the polypeptide to lower levels than detected, i.e., to decrease the effectiveness of the agent.

C. Methods of Treatment

The present invention provides for both prophylactic and therapeutic methods of treating a subject at risk of (or susceptible to) a disorder or having a disorder associated with aberrant expression or activity of a polypeptide of the invention and/or in which the polypeptide of the invention is involved. Disorders characterized by aberrant expression or activity of the polypeptides of the invention are described elsewhere in this disclosure.

1. Prophylactic Methods

5

10

15

20

25

30

In one aspect, the invention provides a method for preventing in a subject, a disease or condition associated with an aberrant expression or activity of a polypeptide of the invention, by administering to the subject an agent which modulates expression or at least one activity of the polypeptide. Subjects at risk for a disease which is caused or contributed to by aberrant expression or activity of a polypeptide of the invention can be identified by, for example, any or a combination of diagnostic or prognostic assays as described herein. Administration of a prophylactic agent can occur prior to the manifestation of symptoms characteristic of the aberrance, such that a disease or disorder is prevented or, alternatively, delayed in its progression. Depending on the type of aberrance, for example, an agonist or antagonist agent can be used for treating the subject. The appropriate agent can be determined based on screening assays described herein.

2. Therapeutic Methods

Another aspect of the invention pertains to methods of modulating expression or activity of a polypeptide of the invention for therapeutic purposes. The modulatory method of the invention involves contacting a cell with an agent that modulates one or more of the activities of the polypeptide. An agent that modulates activity can be an agent as described herein, such as a nucleic acid or a protein, a naturally-occurring cognate ligand of the polypeptide, a peptide, a peptidomimetic, or other small molecule. In one embodiment, the agent stimulates one or more of the biological activities of the polypeptide. Examples of such stimulatory agents include the active polypeptide of the invention and a nucleic acid

molecule encoding the polypeptide of the invention that has been introduced into the cell. In another embodiment, the agent inhibits one or more of the biological activities of the polypeptide of the invention. Examples of such inhibitory agents include antisense nucleic acid molecules and antibodies. These modulatory methods can be performed *in vitro* (e.g., by culturing the cell with the agent) or, alternatively, *in vivo* (e.g., by administering the agent to a subject). As such, the present invention provides methods of treating an individual afflicted with a disease or disorder characterized by aberrant expression or activity of a polypeptide of the invention. In one embodiment, the method involves administering an agent (e.g., an agent identified by a screening assay described herein), or combination of agents that modulates (e.g., up-regulates or down-regulates) expression or activity. In another embodiment, the method involves administering a polypeptide of the invention or a nucleic acid molecule of the invention as therapy to compensate for reduced or aberrant expression or activity of the polypeptide.

Stimulation of activity is desirable in situations in which activity or expression is abnormally low or down-regulated and/or in which increased activity is likely to have a beneficial effect. Conversely, inhibition of activity is desirable in situations in which activity or expression is abnormally high or up-regulated and/or in which decreased activity is likely to have a beneficial effect.

The contents of all references, patents, and published patent applications cited throughout this application are hereby incorporated by reference.

Deposit of Clones

5

10

15

20

25

30

Each of these deposits was made merely as a convenience to those of skill in the art. These deposits are not an admission that a deposit is required under 35 U.S.C. §112.

Clone EpT202, encoding human TANGO 202 was deposited with the American Type Culture Collection (ATCC[®], 10801 University Boulevard, Manassas, VA 20110-2209) on April 21, 1999 and was assigned Accession Number 207219. This deposit will be maintained under the terms of the Budapest Treaty on the International Recognition of the Deposit of Microorganisms for the Purposes of Patent Procedure.

Clone EpTm202, encoding murine TANGO 202 was

deposited with ATCC® on April 21, 1999 and was assigned (composite) Accession Number 207221. This deposit will be maintained under the terms of the Budapest Treaty on the International Recognition of the Deposit of Microorganisms for the Purposes of Patent Procedure.

Clone EpT234, encoding human TANGO 234 was deposited with ATCC[®] on April 2, 1999 and was assigned Accession Number 207184. This deposit will be maintained under the terms of the Budapest Treaty on the International Recognition of the Deposit of Microorganisms for the Purposes of Patent Procedure.

5

10

15

20

25

30

Clone EpT265, encoding human TANGO 265 was deposited with ATCC[®] on April 28, 1999 and was assigned Accession Number 207228. This deposit will be maintained under the terms of the Budapest Treaty on the International Recognition of the Deposit of Microorganisms for the Purposes of Patent Procedure.

Clone EpT273, encoding human TANGO 273 was deposited with ATCC® on April 2, 1999 and was assigned Accession Number 207185. This deposit will be maintained under the terms of the Budapest Treaty on the International Recognition of the Deposit of Microorganisms for the Purposes of Patent Procedure.

Clone EpTm273, encoding murine TANGO 273 was deposited with ATCC® on April 2, 1999 and was assigned (composite) Accession Number 207221. This deposit will be maintained under the terms of the Budapest Treaty on the International Recognition of the Deposit of Microorganisms for the Purposes of Patent Procedure.

Clone EpT286, encoding human TANGO 286 was deposited with ATCC[®] on April 20, 1999 and was assigned (composite) Accession Number 207220. This deposit will be maintained under the terms of the Budapest Treaty on the International Recognition of the Deposit of Microorganisms for the Purposes of Patent Procedure.

Clone EpT294, encoding human TANGO 294 was deposited with ATCC[®] on April 20, 1999 and was assigned (composite) Accession Number 207220. This deposit will be maintained under the terms of the Budapest Treaty on

the International Recognition of the Deposit of Microorganisms for the Purposes of Patent Procedure.

Clone EpT296, encoding human INTERCEPT 296 was deposited with ATCC® on April 20, 1999 and was assigned (composite) Accession Number 207220. This deposit will be maintained under the terms of the Budapest Treaty on the International Recognition of the Deposit of Microorganisms for the Purposes of Patent Procedure.

5

10

15

20

25

30

Clones containing cDNA molecules encoding human TANGO 286, human TANGO 294, and INTERCEPT 296 were deposited with ATCC[®] on April 21, 1999 as Accession Number 207220, as part of a composite deposit representing a mixture of five strains, each carrying one recombinant plasmid harboring a particular cDNA clone.

To distinguish the strains and isolate a strain harboring a particular cDNA clone, an aliquot of the mixture is streaked out to single colonies on nutrient medium (e.g., LB plates) supplemented with 100 mg/ml ampicillin, single colonies are grown, and then plasmid DNA is extracted using a standard mini-preparation procedure. Next, a sample of the DNA mini-preparation is digested with a combination of the restriction enzymes Sall, Not1, and DraII and the resulting products are resolved on a 0.8% agarose gel using standard DNA electrophoresis conditions. This digestion procedure liberates fragments as follows:

- 1. human TANGO 286 (clone EpT286): 1.85 kB and .1 kB (human TANGO 286 has a *DraII* cut site at about base pair 1856).
- 2. human TANGO 294 (clone EpT294): 1.4 kB and .6 kB (human TANGO 294 has a *Dra*II cut site at about base pair 1447).
- 3. human INTERCEPT 296 (clone EpT296): .4 kB, 1.6 kB, and .1 kB (human INTERCEPT 296 has *Dra*II cut sites at about base pair 410 and at about base pair 1933).

The identity of the strains can be inferred from the fragments liberated.

Clones containing cDNA molecules encoding mouse TANGO 202 and mouse TANGO 273 were deposited with ATCC® on April 21, 1999 and were assigned Accession Number 207221, as part of a composite deposit representing a mixture of five strains, each carrying one recombinant plasmid harboring a

particular cDNA clone. To distinguish the strains and isolate a strain harboring a particular cDNA clone, an aliquot of the mixture is streaked out to single colonies on nutrient medium (e.g., LB plates) supplemented with 100 mg/ml ampicillin, single colonies are grown, and then plasmid DNA is extracted using a standard mini-preparation procedure. Next, a sample of the DNA mini-preparation is digested with a combination of the restriction enzymes Sal I, Not I, and Apa I, and the resultant products are resolved on a 0.8% agarose gel using standard DNA electrophoresis conditions. This digestion procedure liberates fragments as follows:

5

10

15

20

25

30

- 1. mouse TANGO 202 (clone EpTm202): 3.5 kB and 1.4 kB (mouse TANGO 202 has a *Apa* I cut site at about base pair 3519).
- 2. mouse TANGO 273 (clone EpTm273): .3 kB and 2.6 kB (mouse TANGO 273 has a Åpa I cut site at about base pair 298).

The identity of the strains can be inferred from the fragments liberated.

Human TANGO 202, human TANGO 234, human TANGO 265, and human TANGO 273 were each deposited as single deposits. Their clone names, deposit dates, and accession numbers are as follows:

- human TANGO 202: clone EpT202 was deposited with ATCC[®] on April 21, 1999, and was assigned Accession Number 207219.
- human TANGO 234: clone EpT234 was deposited with ATCC[®] on April 2, 1999, and was assigned Accession Number 207184.
- 3. human TANGO 265: clone EpT265 was deposited with ATCC® on April 28, 1999, and was assigned Accession Number 207228.
- 4. human TANGO 273: clone EpT273 was deposited with ATCC[®] on April 2, 1999, and was assigned Accession Number 207185.

All publications, patents, and patent applications referenced in this specification are incorporated by reference into the specification to the same extent as if each individual publication, patent, or patent application had been specifically and individually indicated to be incorporated herein by reference.

Equivalents

5

Those skilled in the art will recognize, or be able to ascertain using no more than routine experimentation, many equivalents to the specific embodiments of the invention described herein. Such equivalents are intended to be encompassed by the following claims.

What is claimed is:

- 1. An isolated nucleic acid molecule selected from the group consisting of:
- a) a nucleic acid molecule having a nucleotide sequence which is at least 40% identical to the nucleotide sequence of any of SEQ ID NOs: 1, 2, 9, 10, 17, 18, 25, 26, 33, 34, 45, 46, 53, 54, 67, 68, 72, and 73, the nucleotide sequence of a cDNA of a clone deposited as one of ATCC[®] 207219, 207184, 207228, 207185, 207220, and 207221, or a complement thereof;
- b) a nucleic acid molecule comprising at least 15 nucleotide residues and having a nucleotide sequence identical to at least 15 consecutive nucleotide residues of any of SEQ ID NOs: 1, 2, 9, 10, 17, 18, 25, 26, 33, 34, 45, 46, 53, 54, 67, 68, 72, and 73, the nucleotide sequence of a cDNA of a clone deposited as one of ATCC[©] 207219, 207184, 207228, 207185, 207220, and 207221, or a complement thereof;
- c) a nucleic acid molecule which encodes a polypeptide comprising the amino acid sequence of any of SEQ ID NOs: 3-8, 11-16, 19-24, 27-32, 35-44, 47-52, 55-66, 69, and 74, the amino acid sequence encoded by a cDNA of a clone deposited as one of ATCC[®] 207219, 207184, 207228, 207185, 207220, and 207221, or a complement thereof;
- d) a nucleic acid molecule which encodes a fragment of a polypeptide comprising the amino acid sequence of any of SEQ ID NOs: 3-8, 11-16, 19-24, 27-32, 35-44, 47-52, 55-66, 69, and 74, or the amino acid sequence encoded by a cDNA of a clone deposited as one of ATCC[®] 207219, 207184, 207228, 207185, 207220, and 207221, wherein the fragment comprises at least 8 consecutive amino acid residues of any of SEQ ID NOs: 3-8, 11-16, 19-24, 27-32, 35-44, 47-52, 55-66, 69, and 74, or the amino acid sequence encoded by a cDNA of a clone deposited as one of ATCC[®] 207219, 207184, 207228, 207185, 207220, and 207221; and
- e) a nucleic acid molecule which encodes a naturally occurring allelic variant of a polypeptide comprising the amino acid sequence of any of SEQ ID NOs: 3-8, 11-16, 19-24, 27-32, 35-44, 47-52, 55-66, 69, and 74, wherein the nucleic acid molecule hybridizes with a nucleic acid molecule consisting of the nucleotide sequence of any of SEQ ID NOs: 1, 2, 9, 10, 17, 18, 25, 26, 33, 34, 45, 46, 53, 54, 67, 68, 72, and 73, the nucleotide sequence of a cDNA of a clone deposited as one

of ATCC® 207219, 207184, 207228, 207185, 207220, and 207221, or a complement thereof under stringent conditions.

- 2. The isolated nucleic acid molecule of claim 1, which is selected from the group consisting of:
- a) a nucleic acid having the nucleotide sequence of any of SEQ ID NOs: 1, 2, 9, 10, 17, 18, 25, 26, 33, 34, 45, 46, 53, 54, 67, 68, 72, and 73, the nucleotide sequence of a cDNA of a clone deposited as one of ATCC* 207219, 207184, 207228, 207185, 207220, and 207221, or a complement thereof; and
- b) a nucleic acid molecule which encodes a polypeptide having the amino acid sequence of any of SEQ ID NOs: 3-8, 11-16, 19-24, 27-32, 35-44, 47-52, 55-66, 69, and 74, or the amino acid sequence encoded by a cDNA of a clone deposited as one of ATCC[®] 207219, 207184, 207228, 207185, 207220, and 207221, or a complement thereof.
- 3. The nucleic acid molecule of claim 1, further comprising vector nucleic acid sequences.
- 4. The nucleic acid molecule of claim 1 further comprising nucleic acid sequences encoding a heterologous polypeptide.
 - 5. A host cell which contains the nucleic acid molecule of claim 1.
 - 6. The host cell of claim 5 which is a mammalian host cell.
- 7. A non-human mammalian host cell containing the nucleic acid molecule of claim 1.
 - 8. An isolated polypeptide selected from the group consisting of:
- a) a fragment of a polypeptide comprising the amino acid sequence of any of SEQ ID NOs: 3-8, 11-16, 19-24, 27-32, 35-44, 47-52, 55-66, 69, and 74, or the amino acid sequence encoded by a cDNA of a clone deposited as one of ATCC®

207219, 207184, 207228, 207185, 207220, and 207221, wherein the fragment comprises at least 8 contiguous amino acids of any of SEQ ID NOs: 3-8, 11-16, 19-24, 27-32, 35-44, 47-52, 55-66, 69, and 74, or the amino acid sequence encoded by a cDNA of a clone deposited as one of ATCC[®] 207219, 207184, 207228, 207185, 207220, and 207221;

- b) a naturally occurring allelic variant of a polypeptide comprising the amino acid sequence of any of SEQ ID NOs: 3-8, 11-16, 19-24, 27-32, 35-44, 47-52, 55-66, 69, and 74, or the amino acid sequence encoded by a cDNA of a clone deposited as one of ATCC[®] 207219, 207184, 207228, 207185, 207220, and 207221, wherein the polypeptide is encoded by a nucleic acid molecule which hybridizes to a nucleic acid molecule consisting of the nucleotide sequence of any of SEQ ID NOs: 1, 2, 9, 10, 17, 18, 25, 26, 33, 34, 45, 46, 53, 54, 67, 68, 72, and 73, the nucleotide sequence of a cDNA of a clone deposited as one of ATCC[®] 207219, 207184, 207228, 207185, 207220, and 207221, or a complement thereof under stringent conditions; and
- c) a polypeptide which is encoded by a nucleic acid molecule comprising a nucleotide sequence which is at least 40% identical to a nucleic acid consisting of the nucleotide sequence of any of SEQ ID NOs: 1, 2, 9, 10, 17, 18, 25, 26, 33, 34, 45, 46, 53, 54, 67, 68, 72, and 73, the nucleotide sequence of a cDNA of a clone deposited as one of ATCC[©] 207219, 207184, 207228, 207185, 207220, and 207221, or a complement thereof.
- 9. The isolated polypeptide of claim 8 having the amino acid sequence of any of SEQ ID NOs: 3-8, 11-16, 19-24, 27-32, 35-44, 47-52, 55-66, 69, and 74, or the amino acid sequence encoded by a cDNA of a clone deposited as one of ATCC® 207219, 207184, 207228, 207185, 207220, and 207221, or a complement thereof.
- 10. The polypeptide of claim 8, wherein the amino acid sequence of the polypeptide further comprises heterologous amino acid residues.

11. An antibody which selectively binds with the polypeptide of claim 8.

- 12. A method for producing a polypeptide selected from the group consisting of:
- a) a polypeptide having an amino acid sequence comprising any of SEQ ID NOs: 3-8, 11-16, 19-24, 27-32, 35-44, 47-52, 55-66, 69, and 74, or the amino acid sequence encoded by a cDNA of a clone deposited as one of ATCC[®] 207219, 207184, 207228, 207185, 207220, and 207221, or a complement thereof;
- b) a polypeptide comprising a fragment of a protein having the amino acid sequence of any of SEQ ID NOs: 3-8, 11-16, 19-24, 27-32, 35-44, 47-52, 55-66, 69, and 74, or the amino acid sequence encoded by a cDNA of a clone deposited as one of ATCC® 207219, 207184, 207228, 207185, 207220, and 207221, or a complement thereof, wherein the fragment comprises at least 8 contiguous amino acid residues of any of SEQ ID NOs: 3-8, 11-16, 19-24, 27-32, 35-44, 47-52, 55-66, 69, and 74, or the amino acid sequence encoded by a cDNA of a clone deposited as one of ATCC® 207219, 207184, 207228, 207185, 207220, and 207221, or a complement thereof; and
- c) a naturally occurring allelic variant of a polypeptide having an amino acid sequence comprising the sequence of any of SEQ ID NOs: 3-8, 11-16, 19-24, 27-32, 35-44, 47-52, 55-66, 69, and 74, or the amino acid sequence encoded by a cDNA of a clone deposited as one of ATCC® 207219, 207184, 207228, 207185, 207220, and 207221, or a complement thereof, wherein the polypeptide is encoded by a nucleic acid molecule which hybridizes with a nucleic acid molecule consisting of the nucleotide sequence of any of SEQ ID NOs: 1, 2, 9, 10, 17, 18, 25, 26, 33, 34, 45, 46, 53, 54, 67, 68, 72, and 73, the nucleotide sequence of a cDNA of a clone deposited as one of ATCC® 207219, 207184, 207228, 207185, 207220, and 207221, or a complement thereof under stringent conditions;

the method comprising culturing the host cell of claim 5 under conditions in which the nucleic acid molecule is expressed.

13. A method for detecting the presence of a polypeptide of claim 8 in a sample, comprising:

- a) contacting the sample with a compound which selectively binds with a polypeptide of claim 8; and
- b) determining whether the compound binds with the polypeptide in the sample.
- 14. The method of claim 13, wherein the compound which binds with the polypeptide is an antibody.
- 15. A kit comprising a compound which selectively binds with a polypeptide of claim 8 and instructions for use.
- 16. A method for detecting the presence of a nucleic acid molecule of claim 1 in a sample, comprising the steps of:
- a) contacting the sample with a nucleic acid probe or primer which selectively hybridizes with the nucleic acid molecule; and
- b) determining whether the nucleic acid probe or primer binds with a nucleic acid molecule in the sample.
- 17. The method of claim 16, wherein the sample comprises mRNA molecules and is contacted with a nucleic acid probe.
- 18. A kit comprising a compound which selectively hybridizes with a nucleic acid molecule of claim 1 and instructions for use.
- 19. A method for identifying a compound which binds with a polypeptide of claim 8, the method comprising the steps of:
- a) contacting a polypeptide, or a cell expressing a polypeptide of claim 8 with a test compound; and
 - b) determining whether the polypeptide binds with the test compound.

20. The method of claim 19, wherein the binding of the test compound with the polypeptide is detected by a method selected from the group consisting of:

- a) detection of binding by direct detecting of test compound/polypeptide binding;
 - b) detection of binding using a competition binding assay;
- c) detection of binding using an assay for an activity characteristic of the polypeptide.
- 21. A method for modulating the activity of a polypeptide of claim 8 comprising contacting the polypeptide or a cell expressing the polypeptide with a compound which binds with the polypeptide in a sufficient concentration to modulate the activity of the polypeptide.
- 22. A method for identifying a compound which modulates the activity of a polypeptide of claim 8, comprising:
 - a) contacting the polypeptide with a test compound; and
- b) determining the effect of the test compound on the activity of the polypeptide to thereby identify a compound which modulates the activity of the polypeptide.
- 23. An antibody substance which selectively binds to the polypeptide of claim 8, wherein the antibody substance is made by providing the polypeptide to an immunocompetent vertebrate and thereafter harvesting blood or serum from the vertebrate.

131 426 151 486 171 546 91 306 111 366 11 66 51 186 71 246 31 126 rg c CCA CHC CHC GGA GGA SAGT ng TgT LCTA CIC Y N AAC ATC F P CCC A GCA TACT D GAT н CIG N AAC AAG G GA SGC T ACA င် အ င TGC PCCA ט **₽** ¥ TGG CIC Y N AAT TACT H A GCT r CIC N AAC CCA R AGA Y GAT CAA 299 Y ŋ GGC GC V GTC CCC CAG ၁ ညီ K AAG I ATA မ ၁၅ H A GCC T ACA CAG G GGT Y ACC s AGC Y TAT S Ħ A GCC ၁ ညီ CIC E GAG CCL G GGA F AAC DGAT Z GCC CCC M ATG A GCG R AGG CAC E GAG ဝ ၁၅ AAA TACT × P CCG က္က ည Y TAT E GAG E GAG CAC CIT AAC ი GGG Z A GCG DGAT GGT E GAG N AAC TCC A GCT R CGG N AAC ഗ A GCG TGG r CTG GCA T ACG **&** G GGA F TTT GTCGACCCACGCGTCCGCCCCACGCGTCCGGCCC K AAG A GCG G GGT FTT မ ဗိ V GTG CCT AAA × Y TAT F CIG CIG ი ი M ATG N AAT AGT ഗ ၁ ၁၅ CAG ACC R AGG T ACG A GCC ng TgT E GAG ᆮ CIC ∓GG TGG ၁ ညီ 286 CAG T ACA ი მმმ P CCA Ö A GCG A GCT ACT FEC FEC K AAG N AAC CCC SAGT R CGG A GCG c TGT ი ი GGC CGC S AGC CCT CTA Н E GAG ပ ပို့ Y V GTG I ATA C TGT **A** 600 PCCT F TTT D GAC E CAA K ÄAA CCT S TCC G P

Fig. 1.

2/96

231 726 271 846 191 606 291 906 211 666 E GAG D GAC PCCT A GCC N AAT E GAG DGAT PCCT G G G G V GTG PCCA IATC P M ATG ස වි S ICC R CGC O CAG င ၁ရိုင GEC Y FAT V GTT DGAC S AGC DGAT PCCA I ATC V GTC A GCG R AGG L E GAA R CGG S TCT G GGG IATC V GTG IATC S TCG F E GAA R AGG s TCT T ACC DGAC H E FTTC ရ ၁၉ W TGG R AGG F. TTC Y TAT K AAG S TCT D M ATG Y IATC L TTG V GTC R CGC E GAG ပ္သင္မ DGAC IATC **₽** g 660 A A GCC G GGT ი ემც S TCA V GTC F TTT CAA LCIA V GTC Y Y R AGG V GTC Y C TGT LCTA F CCC N AAC ი მმმ D GAC L TTA CCC R G GGG L ACG F TTC H V GTT CAA T ACC A GCC S AGC T ACC GGT S SCT GCT CAC င် အ Y TAT F Y V GTC F CCT D GAT ₽ 9 9 9 9 T ACC H 9 9 9 9 N AAC G GGA I ATC ი მიც ပ္ပင္ပ D GAC D GAT F TTC Q CAG AAC AAC H FITC GCC CCC L V GTG န ICC A GCC F ပ္ပဋ s. Icc CTT LCTG CHC O S S

Fig. 1B

S AGT

CIC

N AAC

A GCC

O CAG

GAG

T ACG

IATC

V GTG

E GAG

A

T ACG

N AAC

A GCT E

V GTT

R AGA

H

s AGC

G GGA

A GCT

ი მმმ

M ATG

CCC CCC

PCCA

A GCG

¥ TGG

s TCC

N AAT

s AGC

G GGT

PCCA

V GTC

T ACT GGA

K AAG D GAT S

PCCA

CAC

s AGC

PCCC

SAGC

T ACC

TACC

IATC

V GTC

Y TAT

CHC

V GTC

K AAA

S TCC

S

R

A GCC

AGCT

3/96 476 431 1326 451 1386 A GCA V GTA L S TCC I ATT DGAC PCCT R CGC ი წნმ GAC A GCC K AAG Y DGAT T ACA S TCA V GTC F TTT A GCT CAA T ACA IATT P L SAGC V GTT I AŤC OCAG R CGT ₩ TGG IATC LCTC H G GGT S E GAA K AAG CTC ი გიგ CHC ACT K AAA F K AAA A GCA s TCG K AAG L CTG T ACA T ACT V GTC G GGT ი მვვ H · Y TAT PCCA F TTT V GTC L CAA I ATC T ACA L H s TCC S AGT W TGG I ATA C TGT I ATT

Fig. 1C

4/96

1620 1657 1540 AAACCCCACTGTGCCTAGGACTTGAGGTCCCTCTTTGAGCTCAAGGCTGCCGTGGTCAACCTCTCTGTGGTTCTTCTCTC TGACAGACTCTTCCCTCTCTCCTCTGCCTCGGCCTCTTCGGGGAAACCCTCCTACAGACTAGGAAGAGGCCACCT GCTGCCAGGCAGGCAGAGCCTGGATTCCTCCTGCTT

Fig. 11

39 99 377 79 137 197 GAG GAT NAAC 9 囝 929 GCA ¥ ₹ L CTG GGT G GGA CTG AAC GGA ပ္ပင္ပ n TGT E GAA CCA ი მმმ TIC K AAG N AAC CCT IGC မှ ည AGC ი გვი GIG GAG G GGT Y TAC GAC K AAG CCC Ω CAA L CTG gga CIG GGA **4** 9 ACG နှင့် မြင့်င GAT င္ပင္ပင T ACA N AAC CCA ₩ TGG CCC Y AAT Z ი 667 SAGC CCG G AGA Q CAG CCC TGC TGC H Ö TACA ₽ GCG TAT CAG GGA က္သ F TTC AAT CGG AGG CAC TACT C ATG

Fig. 1F

199 219 737 239 259 857 139 497 179 617 V GTC L S A GCT G GGG GGT S TCA F TTT သ မြင့်င Y R AGA V GGA ACC H L F TTT P K AAG G P N AAC 900 T ACC နှ ပို AAA AAA K. AAG 9 9 9 9 TACT A T G CAG FITC H SAGT FTTC W O CAG ACC Y TAC ACG G GGT A GCC N AAC ACC R AGA င် မြင့်င O CAG H င TGC Y TAC F ဗ ဗ္ဗ DGAC **A** 500 T ACG DGAC A GCC ACC H ဗ္ဗဗ္ဗ SAGT PCCT CIC ж С<u></u>С N AAT ი მმმ ဗ္ဗဗ္ဗ D 3AC I ATC GAC I ATT P C TGT N AAC FITC V GTG P ့ မ L L ာ်ရှိ E FTC G GGG CIC F STCT PCCT V GTC D GAC E GAG CIGI S AGC C TGT TACT A GCC PCCA Y S AGT D GAC CCT G GGA AAC IATC F TTC PCCA K AAG CIGI N AAT e Tit S TCC GGA ာ ဦင် **A** 3000 ပရိုင္ပ Ý TAC ₩ IGG ACC r CFC GAC H VGTT V GTG Y CAA Y TAT EGAG I ATC R CGG A 3CA D GAT а 66С STC I ATA ACC V GTG I ATC S K AAG I ATT ACC. GGA Y ACC s TCA s AGC R AGG A GCC D 3AC DGAC E GAG A ၁၉ A GCA ာ ဦင် CIC W AGG E GAG ဗ္ဗဗ္ဗ K AAG MATG A GCG D GAC M ATG Y TAC ATC EGAG ₽ 200 ာ ဦင် N AAC ი მემ ი მი H L

Fig. 1F

419 439 1397 399 1277 319 1037 R AGG V GTG R S TCC CIC I ATA GTC P CC CC EGAG STCA N AAT EGAG Y IAT H CHC ი მმმ R CGA L TTA L S AGC L ITG A GCA R AGA PCCT H GG H CAT LCTT L CHG V GTG V GTC T ACC A GCC S TCA G GGG R AGA S TCT K AAA I ATT A M ATG L L CAG s TCC A GCC T ACG V GTC F TTT C TGC s TCT N AAC STCC V GTA CTGT S AGC G GGA L CIG F V GTC P CCA V GTC H Q CAG OCAG A GCC A GCC A GCC F TTC V GTC S TCT S TCT A GCT GGA CAA DGAT R AGA TACT PCCT A GCT ACT L CTG CAG G GGG N AAT s AGC Q CAG K AAG CAA LCTA R AGA PCCT E GAG V GTC PCCG T GG s TCG L TTA P CCA P CCG I ATC T ACC CGC CGC CGC CAG S AGT PCCT E GAG L TTG LCTT E GAG L CAG N AAC . D GAT PCCA H T ACA O CAG Q CAG M ATG P N AAC s AGC A GCC S TCT A GCC CCC H FTTC E ACA S STCT s AGC EGAG CAA PCCA Q CAG L FTTC FTT E GAG PCCC ი მიც S TCA နှ ၂၄၄ Y IAT K AAG CCA B L ACC. TACC ACC. L TTG S EGAG G GGA CAG I ATC ATC · S , I ATC A GCC I ATT

Fig. 1(

1493

V R V N K M T A I P S * GTC AGA GTC AAG ATG ACC GCA ATC CCC TCG TGA

.2283 2599 2836 2362 2520 2678 2204 2441 TAGCCCTCAAGTAGTTGCCAATCCTGTGGAATCAGAATTCAGCCTGTCTTCCTGTCTCTCAGCCCAAGCCTGTAGCCTAG AGCTGGGGCTGTAGCCTAGAGCTGGGGCTGTAGCCTTGGGGGCTGTAGCACAGAGCTGGGGGCTGTAGCCTAGAGC TGGGGCTGTAGCACAGAGCTGGGGCTGTAGCCTAGAGCTGGGGCTGTAGCACAGAGCTGGGGCTGTAGCACAGAGCTGG TTCCTACGTGAGGTGTCATCATTTTAAAAGCAGATCAAAACTACCGCGAGTTTTGTCCTTTGTCCCTTATCATGGGAGC AGAGTAGGAGTAAGGGCTCTGGTCTTGCTCATTGTCCCCCAGACAGGGAGGCAGGAAAAGGTCAGGCTTGGGAACTGGA CTGCAGTCTGGAAGTGGCCTTTTGTCAGCAGCTGTGCCCTGAAGGTAGACCTTGGTCACTCTCCTGCCAGCCCTTGA CCCTGTCTTTACAGTTTGCAATAGAGCCAGACTGAAAGAACTGTCAGGTTTTCTAGGCTGGCCTGGTTCCCCACTAAGA GATCCTCCCAGGAAAAGCTGCAAGATTGAGAGCCCAGCTGCAGTTGGGAGAGAGGGCCCATCCCCGACTGAGAAGTC GCCTCTGCTCTCCTGGGTACCCTCCTGGAACACCATGCTAACCTTCCCCGAGTCTCTCAGTCACTGCCATTGAGGCCTC TCCTCTAGCTGCTGCTCCCCAGGACTGTCTGGGGCCATCTGGGGATCAGGGAGAGGGCAGCAGCAGAGTACTGACGAGGCAG TACAGGGGTACTAAGCTAGGGGGTCATCATCTTCATTTGATCTGGGAAAGGCTACAGGCTCCTGGATGTGAAGACAGGCC GTACCAGCCTGGCTGCTCTGCTGGGGATGGTAAGACAGGCCCAGGCTGACAGGACACAGCTGGACCTGACTCCAGAAGA **AGTCCGAGGGGACTGAGAGCAGGGCCACACACCAGATGTCATCTTTCTAGAGGGTTCTTTTTAGTACCCACTGACCAATGG** GGCAAGCCTGAGGATTGGTCCATCTGTCCATGGAACAGACACAGTGAACTTCCTGGATACTAGACTTAACTAGCC STGACTGAAGCCCACGCCTGCATGAGAGGCTCCGCTCCAAGCTCGAGTTTGCTCCCCTGAGTTCTCCTCTGATGAGTTC CACTACATAAGAAGACCACTGGAAATAGACTGACAGGAGCAGGTTCCACTCTAGGCTGTCCATAGCGTTTGCAGGACTC CCTGCCTTCCCATTCACCACCATCTCTTTTGGGAGCACCCTGCTTTAGAGGCAGCCCAGCCTGGGATCCTCCATCACAT GTGGCATTGGCGCCCTAGAGGCCCAGAGGCCCAGTGTAGGCTTGGAGCTTTCTCTGCTGCCAACTACCATGTGTCATCT

Fig. 11

3942 3389 3626 3705 3863 4100 4258 4337 4416 4574 3784 4495 3310 4021 ATCTCGGACACAGAAGCAAGCCCCAGTGTGGTGGCAGCTGCGGCTCAGCATTGGTGTCCCCAGGAAGGGCGGTGGATG GCGGCTGCCTAAAGTGAGCAAGAGAACAGAGCTCTGGACTTCTCTAAATGTGGGGCTCTGGCTTCAGACTCCTCAGCCA AAAGCTCTTGAAGATCAAAGCTCTGGCGGGTACAGCTGTCCTGGCCTGTGGGCCAGCCCATGGGATGTGCCTGGGCCAG GTGCCACCCCACGGCTCACTGTCATCCCAGGAGGGACCCCACCTGATGCTCCTCATCATCGCTGGCCTGACACTATCA GAGCTCGCGCCGGCTGTTGCCAGGGACAGACTGACTACACTTGACCTTCAAGAGCACTTAGAAGTGGATGGCCTCCAGA CTCTGTCAGCCTCTGCAGGGGCCACACAAGTCTCCCGAGCCAAGTCCACAAGCCTCCATGGTTCCCTGGCTCCTCTCCT GGCTGCACCCCCACCCTGGTCTGCCAACAGAACCTGGGGGCCTCACACGGGCTCCTGTCTTGCCAAGCTGGAGCTGAGC AGATCCAGCGAGGGAGCTGCCATCCCCGCCACCTTCATAGCAGCAAGACCTTCCCATTTCCAATCTCACCTCCAGCAG GGATATGACTTTGGACAACAAGGCTTTATTTGTAAATATGCTCTTAATATGCAACTTTGAGAATAAGATAGAAACATCA CCCTGAGACCAAGTGTTGAGTCACAGAGTGCCATGTGCGTAGTGCATAAAGGATATGGGTTCTTAACCAGGGAAGGCTC ATAGCAGGCCAGGACATTTTTTCAGCTCAGAGCACTGGCCCCCAGGCTTCCTCTAAGCCACCACTCACCTGTCTTCCTT TGCCCACGCTCCTTTTGCTGTGGGCCTGGCACAGCCCAACACTGCAGGGCCCACCTTCTCTCTTGGGGGGTAGGGACAC GTGGAGTGTCCTGTTTGATGTCTGAGGTCTGCTTTGGGTACCGCCCTGGGAACTGCTAACCTCCGATTGGTCCTTTGT TCCTGTCACACTGCTTACAAAGCAGAGACAGAGTAGGAAAGAGGTCTTCATCCTCTCCCACATCAGCAAGGATAGGGCT TTGTTGTAGAAAAAAAAAAAAAAAAAAAAGGGGGGCGCCGC

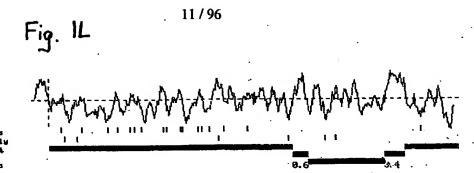
Fig.1

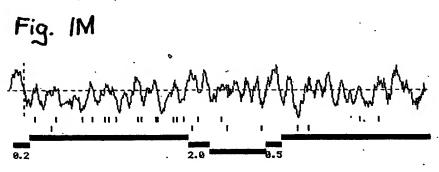
| 70 COHPYNT :::::: | 140 LTGTSKT | 210 SRIILFD | 280 GYTHRV :::::: |
|--|---|--|--|
| 60 KPCLFWNETE KPCLFWNETE | 130 KKDHGNPPPI :::::::::: YKDHGNPPPI 130 | 200 DHTQPCGGDC ::::::::: | 270 DSADMVELLI :::::::: DSADMVELLI 270 |
| 0 | 120 ACQMPGNLGCY :::::::: ACQMPGNLGCY | 190 STECNSVCFGI:::::::::::::::::::::::::::::::::::: | 260 HFSFPLFDIRI :::::::: HFNFTLFDIRI |
| 40 TANGADYRG ::::::::: | 110 SVYWKYCEIP2 :::::::::::::::::::::::::::::::::::: | 180 PDYWKYGEAA ::::::::: PDYWKHGEAA | 250 FIRVPGASHII :::::::::: |
| 30 SPGLGPGPECI : .::: 30 | 90 100 110 120 130 1 VYCRNPDGDVSPWCYVAEHEDGVYWKYCEIPACQMPGNLGCYKDHGNPPPLTGTS::::::::::::::::::::::::::::::::::: | 160 170 180 190 200 2 ISORFKFAGMESGYACFCGNNPDYWKYGEAASTECNSVCFGDHTQPCGGDGRIIL ::::::::::::::::::::::::::::::::::: | 230 240 250 260 270 280 280 270 280 SSVVYSPDFPDTYATGRVCYWTIRVPGASHIHFSFPLFDIRDSADMVELLDGYTHRV::::::::::::::::::::::::::::::::::: |
| 20 ALTLAARPAPSP ::::::::: ALTLAARPAPGP 20 | 90 (CRNPDGDVS) :::::::: (CRNPDGDVS) | 160 SQRFKFAGMES :::::::: SQRFKFAGMES 160 | 230 SVVYSPDFPD: .:::::::: AVVYSPDFPD: 230 |
| 10 20 30 40 50 60 70 70 70 10 70 70 10 70 70 70 70 10 70 70 70 70 70 70 70 70 70 70 70 70 70 | 80 100 110 120 130 140 Hum. LKYPNGEGGLGEHNYCRNPDGDVSPWCYVAEHEDGVYWKYCEIPACQMPGNLGCYKDHGNPPPLTGTSKT ::::::::::::::::::::::::::::::::::: | 150 160 170 180 190 200 210 SNKLTIQTCISFCRSQRFKFAGMESGYACFCGNNPDYWKYGEAASTECNSVCFGDHTQPCGGDGRIILFD :::::::::::::::::::::::::::::::::::: | 220 230 240 250 260 270 280 280 20 20 280 20 20 280 20 20 280 20 20 280 |
| Hum. Mur. | Hum. Mur. | Hum. Mur. | Hum. Mur. |

Fig. 1.

-TEWKD-GLCTAWRPSSSSOSOOLSORFFCM Hum. LARFHGRSRPPLSFNVSLDFVILYFFSDRINQAQGFAVLYQAVKEELPQERPAVNQTVAEVITEQANLSV **LVRLSGRSRPPLSFNVSLDFVILYFFSDRINQAQGFAVLYQATKEEPPQERPAVNQTLAEVITEQANLSV** SAARSSKVLYVITTSPSHPPQTVPGSNSWAPPMGAGSHRVEGWTVYGLATLLILTVTAIVAKILLHVTFK 410 SHRVPASGDLRDCHQPGTSGEIWSIFYKPSTSISIFKKKLKGQSQ-QDDRNPLVSD SHINLIESLHQETLGTVVSLGLLEISGPFSMNLPLQSPSLRRSSRVRVNKMTAIPS 400 330 390 320 SAAHSSKVLYVITPSPSHPPQTAQVAIPGHRQLGPTA 380 310 450 370 440 300 290 430 Mur. Hum. Hum. Mur.

Fig. 1K





Cus Walu out TH

12/96

93 33 126 73 113 366 133 426 FTT GAT CIG a GGA C TGT G GGA SIG L N AAT PCCA L W TGG N AAT A GCC or Grc W TGG TACT C E DGAT L GFC I ATT LCTG r CFG G GGG GGA I ATT R CGG E GAA R AGG ₩ TGG L K AAA IATC H GGT H L TTG CAG Y TAT Q CAG ¥ TGG ည် မြင့် GGA GGA CAA EAG GAG K AAA s TCG TACT GGA H C TGT cTGT V GTA ပည္သိုင္သ N AAC O CAG R AGA E GAA L TTG V GTG V GTG TACT V GTG CAA DGAT F TTC W TGG V GTC V GTG CIC A GCT ACA K AAA G GGT PCCT V GTG G GGA A GCT V GTT r CTG S TCT ACT A GCC N AAT E GAG M ATG F S TCA DGAT CAA s TCA M ATG FITT V GTG A GCC G GA E GAG E GAA L GCGCCCCTCCCGATCTAGAACTAGTA N AAC S AGT T ACA ACT F TTT G GGA N AAT CAG SAGC G GGG ACT RCGT GGA H H IATC s TCT N AAC FTT Y TAT Y TAT C TGT GEC C TGC ₩ TGG M ATG C TGT C TGT TGC F TTT P CCC ი მმმ န TCC A GCC N AAC Ü IGC ပည္သိုင္သ FTTC H GGT .D GAT GTT > R AGA S TCC D GAC GAT S TCT GAT SAGC Ω

Fig. 2A

13/96 253 786 233 726 273 846 293 906 313 966 193 606 213 666 TACT ი მვი N AAT ပ္ပင္ပ Ω G ၁ ညီ V GTA N AAT M ATG N AAC န ၂၄၄ L TTA AGCT V GTG AGCT W TGG F W TGG ng TgT င် သည် V GTG r CIC TACA FTTC PCCT K AAA န ICC **8** H A A A G A 3CC s AGC A 3CA STC L TTG H LCTT Ÿ GTC D GAT N AAC E GAG A N AAT ACT H A GCA G GGT E GAG V GTT E GAG V GTG TACT GGA ၁ ညီ ACC D GAT V GTT N AAT N AAT R AGA N AAT CAC GGT V GTA GGA 뎕다 . G G G L TTG GGA ACC o CAG S AGT V GTA CTGTTGG N AAC s TCT S TCA G GGG GGA V GTA ၁ ၁၅ CTT ာ ဦင် C TGT ¥ TGG s TCT D GAC R AGG W L TTG V GTT နှင့် TCC I ATT L TTA ი მმმ I ATT R AGG Q CAG D GAT H LCTA FTT AAC D GAT N AAC DGAC N AAT GAA 9 9 K AAG STCT S D C IGC ი მგმ DGAT GGA L CAA G GGA S TCT CTGT IATC V GTA S TCA PCCA L W TGG DGAT I ATA SAGT CAG GGA K AAA V GTC V GTG TACT C TGT ₩ TGG RCGT S AGT L DGAT L TTG LCTA ი ი ი GGA I ATT E GAG D GAT AGCT H R AGG LCTA CCC CCC H TGG L ITG RCGC R AGA Y TAT V GTA A GCA PCCT CAA R AGG L TTG ၁ ၂ ၁၅ CTGT R AGA L TTG A GCT G GGT E GAA R AGG

Fig. 2E

14/96 433 1326 493 1506 353 1086 453 1386 TAC P A A A D GAT ¥ TGG G GGA V CHE GCA T ACA n TGT A GCA . K AAA GII Ø G GGA D GAT V GTG C TGT L W TGG I ATT AAG AGCT GAC EGAG A GCA LCTA Y TAT DGAT ₩ TGG CGA œ L CAG CAG R AGA TACA S L TTG N AAT K AAG R AGA R AGG E E GAA ' A GCT င TGC C TGT L TTG C TGT E GAA DGAC I ATT ე ემმ T ACA H ACC VGTT W TGG V GTA Y TAT S AGC D GAT I ATT N AAT C ი წ**GA** A GCA R AGA V GTG S AGT CIC GGA ¥ TGG S TCC V GTG LCIT PCCT A GCT A GCT ССС CCC R AGA GGA . ტ s AGC E GAG D GAC H D GAT A GCC K AAA S TCA D GAT S V GTA EGAG S H H CAA A GCT R AGA R AGA N AAT R AGA A GCT ng T ာ ဦင်င E GAA R CGT ი მვმ N AAT ი მემ R CGA ი მმმ V GTG RCGT Ř. AAG V ACT W TGG V GTG S TCA SAGT FTTC ACT ဗဗ္ဗ ₩ TGG င် TGC C TGC CTT ი მი CHT S TCT CTGT ¥ TGG F STG N AAT N AAC F TTT S TCT TACA R AGG CAG V GTC I ATA R CGA L EGAG S TCT D GAT AAC E GAA N AAC S AGT DGAC SAGC SAGC K AAG DGAC GGA GGA F TTC N AAC A GCA CIG C G GGA

Fig. 2C

613 1866 673 2046 N AAC M ATG S TCG DGAT S AGT IATT ¥ TGG ი მმე ည်း ပည် မြိ C TGT R AGG V GTG ဗ ဗ္ဗဗ I ATT STCC D GAC 8 6 6 7 DGAT GAC I ATC V GTT CIG Y TAT S N AAC N AAT EGAG E GAG SAGC D GAT N AAT T ACC E GAG S TCT DGAT M ATG ·M N AAT R AGA а 66С STCT D GAT G GGA ng TgT V GTG ¥ TGG H а 660 PCCA L G GGA s TCG V GTG T ACA I ATT V GTA ာ Tgo ¥ € ი ი A GCA ပမ္မ L ဗ ၁၅၅ DGAC SAGT C TGT I ATT L K AAA N AAC GAT H S TCT N AAT R AGG W TGG Ω s TCT M. ATG V GTT H L 7 7 7 7 8 O CAG GGA R AGG C TGT K AAG s AGC ပည္ CCT DGAC ရ ၁၅၅ GGA X TAT I ATC K AAG GAT WCAA S TCA GGA n HGT g GGA G GGA L ¥ TGG T ACA F TTT V GTĞ TACA W TGG Y G GGA C TGT G GGA A GCA V GTG S TCT L G GGA I ATT SAGT D GAT V GTG A GCT D GAT V GTT A GCT L ITG GGT E GAG A GCA N AAC D GAT PCCT H S TCA U GAA CAA GGA GAA S TCA L A GCT GGA E GAG U R AGA LCTG SAGT K AAA S n TGC K AAA DGAT C TGT T ACC HCAC G GGA SAGT GGT GGA A GCA DGAC

15/96

Fig. 2D

LCTG

ი მემ M ATG ¥ TGG Q CAG DGAC

N AAT

R AGA

16/96 813 2466 793 2406 833 2526 873 646 753 2286 R AGG E A GCA C L TIA H H LCTG K AAA E IATC ¥ TGG A GCA A GCC V GTG R CGA K AAA V GTG GGA CIT H CIT V GTG I ATA S TCA A GCC CAA L TTA N AAT F TTT H R AGG T ACA C TGT င TGC GAA A GCC H GGT TACT IATC D GAT o CAG ာ TGC R AGA D GAT V GTT E GAA AGCT L TTG V GTT RGT E GAA ¥ TGG H GGA S AGT ი მმმ TACA L SAGT V GTG N AAT V GTT GGA LCTT A GCA S TCT STCT E FTTC S TCT S TCT E GAA LCTT EGAG A GCT A GCC E ာ ဦင် FTTC ng c H M ATG GCC CCC S TCT CAG IATT PCCT E GAA DGAT V GTT M ATG N AAC E GAG ი მმმ N AAT I ATA S TCT FTTC X AAA G GGA MATG R AGA G GGA L TTA D GAT D GAT A GCC K AAG GGA GGA TACT AGCT CTGT E S H D GAT V င TGC c TGT ¥ IGG C TGT GGA V GGA A 200 6 66 66 67 A GCG R AGG ပ္ပင္သ R AGG C TGT V GTT S TCT ¥ TGG TACT s AGC S TCT N AAT N AAT ATC TACT L ж С С A A GCA Q CAG N AAT ₩ TGG L TTA LCTA R AGG ng c STCG CCC TACA GGA s TCT K AAA GAA G GGT

Fig. 21

17/96 993 973 946 E GAG AAT GAC K AAA N AAT CAC DGAT N AAT G GGA GCA E GAG Y TAT GGA GGA CAA E ď GGA ი მვი H r GG L TTA I ATC s GGA ი მმმ PCCA E GAG. L TTA I ATC ACA င TGC င် ပြင် GAC DGAC ¥ TGG T ACC ာ ဦင် C TGT PCCA I ATC V GTA င် TGT E GAA H L TTG Q CAG S TCA H P F TTT R AGA R AGA L CTG A GCT ი გმ s TCC T ACC CHC FTTT PCCT GAC A GCT R AGG SAGT A GCC A GCA PCCA AAA GGC n Igi T ACT H GGA GGA Q CAG မှ ဗို $^{\rm C}$ ტ ი მმმ GGA T ACC E GAG မှ ၁ N AAT L CTG CIT ng. TACT V GTG S TCA C TGT W TGG V GTT LCTG PCCA s AGC S AGC V GTG s AGC D GAC GAC ရ ၁၅၅ T ACA V GTT CIT G GGA M ATG A GCA RCGA W TGG CIC R CGT ი გვ o CAG T ACA V GTT CAA S TCT DGAT GIC H c TGT ဂ ဂရိုင် L TTG V GTG R AGA S AGT H D GGA N AAC ာ ဦင္ပင R AGA I ATC Y TAT .T ACA CAA LCTT LCTA D GAT V GTG L EGAA V GTG P CCA V GTT Y TAT ۳ 2 G GGA L LCTC AAC V GTT STCT DGAC I ATT R CGA CCC S IATC R CGT I ATT LCTT V GTC S TCT R CGG S V GTA E GAG A GCC Y TAT TACT C TGT K AAA

Fig. 2F

18/96 1213 3666 1153 3486 1173 3546 1193 3606 1093 3306 A GCT T ACC s AGC E GAG ¥ TGG V GTC S AGC PCCT C TGT I V GTC S TCC ი მმმ r Tr G G G G C IGI s AGC N AAC V GTT Q CAG I ATC H CAC A GCA FTT I ATT EGAG R AGG R AGA GAC G GGA H s TCC E GAG E GAG TACA R AGG N AAT D GAC R CGA DGAT A GCT K AAG E GAG D GAT EGAG T ACG E GAÀ ၈ ၁၁ S TCT H ი მგმ V GTG ¥ TGG V GTC GGA ACT V GTC R AGG E GAA S AGC W TGG PCCA DGAC T ACG TACA C TGT M ATG A GCC ၁ ၁၅ င TGC SAGT ဗ္ဗဗ္ဗ ი ემე A GCC ¥ TGG N AAC · D GAC Y TAC ¥ TGG L FTTC s TCT ရ ရှင် N AAT G GGT CFG D GAC \mathbf{F} L H LCTC ACC O. CAG CAG R AGG . හ ගිරි R AGG ပည္ပို့ A GCC DGAC S TCT GAT ပ္ပပ္ပ L TTG AAC ၁ ဥ CAG C TGT D ი მმმ ¥ TGG A GCC Y TAT T ACA ¥ TGG V GTG GGA L CTG F K AAG I ATA ACC ი ეწე TACA C IGI ¥ TGG I ATT FTTC STCT ۳ ا STC GTC STCC ရ ၁၅ ၁၅ ပ ၁၅၅ ATC I ¥ TGG L S E GAA E GAA A GCA L TTA CCC CCC PCCT S TCA I. TIG H F K AAG I ATA PCCT ი მ T ACG ဗဗ္ဗ CAA S TCA ဂ TGC C TGC R AGA A GCC A GCC O CAG I. I. ი ^{_}ი მმმ TACA CHC C TGT ი მმმ

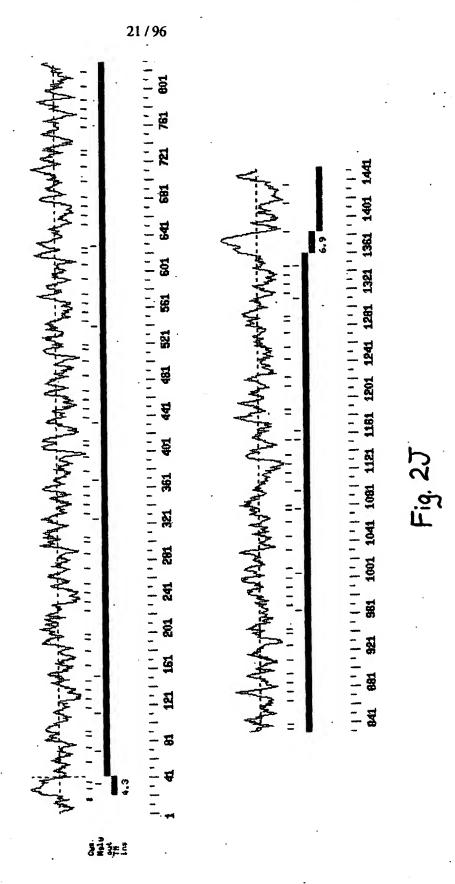
Fig. 2G

19/96 1413 4266 1333 1353 4086 1373 1393 4206 1293 3906 1313 3966 T ACC L CTG M ATG SAGT L r CHG r GFC D GAC CAG E GAG DGAC S TCA CIC G P DGAC D GAT A GCT CIT L M ATG G GGA GGA K AAA C TGT s TCT DGAT E GAG W TGG L 9 9 9 9 L TTG CHC H CAT V GTG ပ္ပင္ပ ¥ TGG S TCG ი მვვ H CAT TACA C TGT CCC CCC AAA RCGI × 9 GGC · I K AAA Q CAG CAA F g G G FTT A G GGA I ATC AAA L TTA ₩ TGG LCTG T ACC × Q CAG N AAT S TCC CAG a GGA H S TCT SAGT ပ ပို့ CAG TACT s TCC VGTT E GAG R AGA CTGT ၁ ၁၅ E GAG C TGT DGAC R AGG L TTA R CGA A GCA GGA GGA D GAT GAA H V GTG CAG ₩ TGG V GTG IATT ည်မှု GG V GTG မ ၁၉ ¥ TGG s TCT ¥ TGG LCTA ရ ဝင္ပင CIT CIGI T ACA I ATC E GAA F TTT A GCT A GCA T ACG G GGT F TTT R AGG E GAG s TCG S TCA LCIC A GCG D GAT L TTA R AGG E GAG EGAG V GTG A GCT E H F TTT R AGA A GAC N AAT K AAG R AGA ACC G GGT LCTA E GAG ი მმი L CTG R AGG G GGA CAC S TCA ACC I ATT s TCA GAA STCT DGAC L CTG K AAA S TCC F TTT GGA ¥ ∏GG **A** ၁ ၁ ၁ C TGT A GCC LCIG V GTT ၁ မှ R AGA R CGG DGAC N AAT E GAG s TCC A GCT V GTT

Fig. 21

| | | | 20 / |
|--|--|--------------|---|
| H G T R T S D D T P N H G 1433 CAT GGG ACA AGA ACC TCA GAT GAC ACC CCC AAC CAT GGT 4326 | S L L G V L P A S E A T K 1453 TCG CTG TTG GGA GTT CTT CCT GCC TCT GAA GCC ACA AAA 4386 | 1454 4389 | CTTTAGACTTCCAGGGCTCACCAGATCAACCTCTAAATATCTTTGAAGGAGACAACAACTTTTAAATGAATAAAGAGGA 4468 AGTCAAGTTGCCCTATGGAAAACTTGTCCAAATAACATTTCTTGAACAATAGGAGAACAGCTAAATTGATAAAGACTGG 4547 TGATAATAAAAAATTGAATTATGTATATCACTGTTAAAAAAAA |
| G GGT | K AAA | | 766A 7766 7766 |
| CAT | T ACA | | AAAGA AAGAC STGGC |
| N AAC | ₽ GCC | | SATAI SATAI ACGC |
| GCC CCC | e gaa | | AAATC AATTC ACGG |
| a Acc | S TCT | | TTTI 3CTA |
| D GAC | A GCC | | CAAC: |
| D GAT | PCCT | | ACAA(GGAG! AAAA! |
| S TCA | L | | GGAG; AATA(AAAA) |
| T ACC | VGTT | | IGAA(SAACI AAAA |
| R AGA | G GGA | | CCTT(CCTT(AAAA) |
| T ACA | TTG | | AATA' |
| ტ ტ ტ | L CTG | | rctal ATAA(STTA |
| H | s TCG | | AACCI |
| PCCA | T ACA | | SATCI FTGT(ATAT(|
| D GAC | D GAC | | ACCA(|
| E GAG | S D AGC GAC | | SCICA IGGAA |
| K R AAG AGA | D A GAT GCT | | CTAC |
| | | | CTTTAGACTTCCAGGGCTCACCAGATCAACCTCTAAATATCTTTGAAGGAGACAACATTTTAAATGAATAAAGAGGA AGTCAAGTTGCCCTATGGAAAACTTGTCCAAATAACATTTCTTGAACAATAGGAGAACAGCTAAATTGATAAAGACTGG TGATAATAAAAAATTGAATTATGTATATCACTGTTAAAAAAAA |
| C L L | C E | | PAGAC PAAG1 |
| င မြင့်င | CTGT | * TGA | CTTI AGTO TGAI |
| | | | |

Fig. 2]



| 70 | OGOWG | QGEWG 50 | WGSHN YRNDG | FISSG: | MGRVE : : : : : : : : : : : : : : : : : : : |
|----|---|---|--|---|---|
| 09 | CSGTVEVKE | CEGRVEVKH | 130 VECQHRE .:.: SDCEHSNIKD 110 | 200 /CRQLGCPSS :::::: CAELGCGKA | 270 LRLVGGTNRCM ::: L-MTNGSSQCE(|
| 50 | GIDLELRLVNGDGPCSGTVEVKFQGQW | -GRHLSIRGLCVLLLGTMVGGQALELRLKDGVHRCEGRVEVKHQGEWG 10 20 30 40 | 110 120 GKIWLDDVSCYGNESALWECQH : :: .: .: :: .:: GPIWLLYTSCEGTESTVSDCEH 90 100 11 | 0 160 170 180 190 200 EANLGLRLVDGNNSCSGRVEVKFQERWGTICDDGWNLNTAAVVCRQLGCPSSFISSG ::::::::::::::::::::::::::::::::::: | 250 270 CSHNEDVTLTCYDSSDLELRLVGGTNRCMGRV : : |
| 40 | FLISSFNGTI | -MVGGQ7 | 110 R-HGKIWLDI : ::: PGLGPIWLL) | 180 EQERWGTICD: : : : : ISGEAWIPVSD: | 250 HDCSHNEDVT: ::::::::::::::::::::::::::::::::::: |
| 30 | TCILLINSC | CVLLLGTMVG | 100 AMFREGOAVT : : : 3-FPGGAYFG | 170 NSCSGRVEVK) :::::: | 240 WNCRHRGWGNI : |
| 20 | CHONLESAV | -HLSLRGL | 90 VVCKQLGCPFSFAN ::::::: VVCRQLGCGAAIG- | 160 LGLRLVDGNI .::. -FVRLAGGD | O 230 DDILCQGNELALWNCRH :.: .: : . KEFRCEGEEPELWVCPR |
| 10 | Hum. MMLPQNSWHIDFGRCCCHQNLFSAVVTCILLLNSCFLISSFNGTDLELRLVNGDGPCSGTVEVKFQGQWG | | 80 100 110 120 130 Hum. TVCDDGWNTTASTVVCKQLGCPFSFAMFRFGQAVTR-HGKIWLDDVSCYGNESALWECQHREWGSHN :: .: .: .: .: .: .: .: .: .: .: WC1 TVDGYRWTLKDASVVCRQLGCGAAIG-FPGGAYFGPGLGPIWLLYTSCEGTESTVSDCEHSNIKDYRNDG 60 100 110 | 140 YHGEDVGVNCYG ::::::: | 210 220 240 250 270 VVNSPAVLRPIWLDDILCQGNELALWNCRHRGWGNHDCSHNEDVTLTCYDSSDLELRLVGGTNRCMGRVE |
| | Hum. | WC1 P | Hum. I WC1 I | Hum. C. WC1 Yl | Hum. V |

| 280 330 340 LKIQGRWGTVCHHKWNNAAADVVCKQLGCGTALHFAGLPHLQSGSDVVWLDGVSCSGNESFLWDCRHSGT ::: .: .: .: .: .: .: .: .: .: . | 260 270 380 300 310 320 350 360 370 380 400 410 VNFDCLHQNDVSVICSDGADLELRLADGSNNCSGRVEVRIHEQWWTICDQNWKNEQALVVCKQLGCPFSV :::::::::::::::::::::::::::::::::::: | 40 450 460 470 480 IWINSISCTGNESALWDCTYDGKAKRTCFRRSDAGVICSDKADLDLRLVGAHSPCY ::::::::::::::::::::::::::::::::::: | 360 510 520 530 540 550 QGEWGTVCHDRWSTRNAAVVCKQLGCGKPMHVFGMTYFKEASGPIWLDDVSCIGNESNIWDCE :::::::::::::::::::::::::::::::::::: |
|--|---|---|---|
| 330 VWLDGVSCSG . ::: ILTARFHCSG | 400 CDQNWKNEQP | 470 AGVICSDKAI ::: CSDSR(| 540 ASGPIWLDDV .::::::: |
| 320 LPHLQSGSDVVV ::: .:: GPHLVEEGDQII | 300 390 VRIHEQWWTI | 460 AKRTCFRRSD | 530 HVFGMIYFKE : |
| 310 WNNAAADVVCKOLGCGTALHFAGLPHL WSLANANVICROLGCGVAISTPGGPHL | 380 38NNCSGRVE | 440 CTGNESALWDCTYDGK :: : | 520 CKQLGCGKPM :.::::: CRQLGCGEAL |
| 300 LAADVVCKÖLG | 280 370 3ADLELRLADO 3NQI | 440 ASISCTGNES | 510 KWSTRNAAVV : .::: |
| 290 TVCHHKWNNF : : : : | 260 270 350 360 3 VNFDCLHQNDVSVICSDGADLE ::::::::::::: GGPDCSHGNTASVICS-GNQI- | 330 340 0 430 RRAKPSNEARDIWIN SQPIGSA | 360 500 GEWGTVCHDF : ::::: |
| | | FGS 1 | 490 GRLEVKY ::.:. GRVEILD 390 |
| Hum. WC1 | Hum. WC1 | Hum. WC1 | Hum. WC1 |

Fig. 2I

| | | 560 | 570 | 580 | 590 | 009 | 610 | 620 |
|--------|------|-------------------|--|-------------------|--|---|-------------------|---|
| | Hum. | HSGWGKHNC | HSGWGKHNCVHREDVIVTCSGDATWGLRLVGGSNRCSGRLEVYFQGRWGTVCDDGWNSKAAAVVCSQLDC | DATWGLRLVG | GSNRCSGRLE | VYFQGRWGTV | CDDGWNSKAA | AVVCSQLDC |
| | | • | ••••••••••••••••••••••••••••••••••••••• | | | ••••••••••••••••••••••••••••••••••••••• | • | ••••••••••••••••••••••••••••••••••••••• |
| | WC1 | SRGWGQHNCI | SRGWGQHNCRHKQDAGVICSEFLALRMVSEDQQCAGWLEVFYNGTWGSVCRNPMEDITVSTICRQLGC | EFLALRMVS | SEDQCCAGWLE | VEYNGTWGSV | CRNPMEDITV 510 | STICROLGC |
| | | 460 | 0 / 4 | φ |) N |)) |) 1 1 | |
| | Hum. | 630 PSSIIGMGL(| 630 640 650 660 650 990 PSSIIGMGLGNASTGYGKIWLDDVSCDGDESDLWSCRNSGWGNNDCSHSEDVGVICSDASDMELRLVGGS | 650 DDVSCDGDES | 660 SDLWSCRNSGW | 6/U GNNDCSHSED | ogu VGVICSDASI | 690 MELRLVGGS |
| | WC1 | GDSGTLNSS | Ϋ́ | OWVDRIQCRKTD | DISLWQCPSDPWNYNSCSPKEEAYIWCADSF | NYNSCSPKEE | AYIWCADSR- | ::::. QIRLVDGG 590 |
| ·: | | 530 | 0.44 C |) () |) 0 | 2 |))) (| |
| ,÷, | | 700 | 710 | 720 | . 730 | 740 | 750 | /60 |
| ٠.,٠ | Hum. | SRCAGKVEVI | SRCAGKVEVNVQGAVGILCANGWGMNIAEVVCRQLECGSAIRVSREPHFTERTIHIIMSNSGCTGGEASL | NGWGMNIAEV | /CRQLECGSAI | RVSREPHFTE | RTLHILMSNS | SGCTGGEASL |
| | | • | | • | | • | | |
| . • • | WC1 | GRCSGRVEILDQG | LDQGSWGTICDI 610 | ORWDLDDARV 620 | SWGTICDDRWDLDDARVVCKQLGCGEALDATVSSFFGTGSGF1WLDEVNCKGEESQV 610 620 660 | DATVSSFFGT 640 | GSGPIWLUEN 650 | NCKGEESQV 660 |
| ·.• ., | , | | 9 |) |))) | | | |
| | | 770 | 780 | 790 | 800 | 810 | 820 | 830 |
| *:• | Hum. | WDCIRWEWK | WDCIRWEWKQTACHLNMEASLICSAHRQPRLVGADMPCSGRVEVKHADTWRSVCDSDFSLHAANVLCREL | LICSAHRQPRI | LVGADMPCSGR | VEVKHADTWR | SVCDSDFSL | IAANVLCREL |
| | | ••• | | ••• | | : | ••• | |
| ia. | WC1 | WRCPSWGWR | WC1 WRCPSWGWRQHNCNHQEDAGVICSGFVRLAGGDGPCSGRVEVHSGEAWTPVSDGNFTLPTAQVICAEL | VICSGEVR | LAGGDGPCSGR | VEVHSGEAWT | PVSDGNFTL | PTAQVICAEL |
| • | | 670 | 089 | 069 | 700 | 710 | 720 | 730 |

Fig. 2IV

Hum.

WC1

096.

GCGKAVSVLGHMPFRESDGQVWAEEFRCDGGEPELWSCPRVPCPGGTCLHSGAAQVVCSVYTEVQLMKNG

NCGDAISLSVGDHFGKGNGLTWAEKFQCEGSETHLALCPIVQHPEDTCIHSREVGVVCSRYTDVRLV-NG

LWSCPVTALGGPDCSHGNTASVICSGNHTQVLPQCNDFLSQPAGSAASEESSPYCSDSRQLRLVDGGGPC GGRVEILDQGSWGTICDDDWDLDDARVVCRQLGCGEALNATGSAHFGAGSGPIWLDDLNCTGKESHVWRC LDNCQMTVLGAPPCIHGNTVSVICTGSLTQPLFPCLANVSDPYLSAVPEGSALICLEDKRLRLVDGDSRC AGRVEIYHDGFWGTICDDGWDLSDAHVVCQKLGCGVAFNATVSAHFGEGSGPIWLDDLNCTGTESHLWQC TSQCEGQVEMKISGRWRALCASHWSLANANVVCRQLGCGVAISTPRGPHLVEGGDQISTAQFHCSGAESF KSQCDGQVEINVLGHWGSLCDTHWDPEDARVLCRQLSCGTALSTTGGKYIGERSVRVWGHRFHCLGNESL WC1 WC1 WC1 Hum. Hum.

| 1120 1130 1180 PSRGWGQHDCRHKEDAGVICSEFTALRLYSETETESCAGRLEVFYNGTWGSVGRRNITTAIAGIVCRQLG :::::::::::::::::::::::::::::::::::: | 1190 1200 1210 1220 1230 1240 CGENGVVSLAPLSKTGSGFMWVDDIQCPKTHISIWQCLSAPWERRISSPAEETWITCEDR ::.:::::::::::::::::::::::::::::: | 1250 1260 1270 1280 1390 1300 umIRVRGGDTECSGRVEIWHAGSWGTVCDDSWDLAEAEVVCQQLGCGSALAALRDASFGQGTGTIW .:::::::::::::::::::::::::::::::::::: | 1310 1320 1330 1340 1360 1360 1360 1360 1360 um. LDDMRCKGNESFLWDCHAKPWGQSDCGHKEDAGVRCSGQSLKSLNASSGHLALI ::::::::::::::::::::::::::::::::::: |
|---|--|--|--|
| 1160 1170 CAGRLEVFYNGTWGSVGRRN:::::::::::: CAGWLEVFYNGTWGSVCRSP | 1190 1200 1210 1220 1230 1240 CGENGVVSLAPLSKTGSGFMWVDDIQCPKTHISIWQCLSAPWERRISSPAEETWITCEDR ::: .: .: ::: ::::::::::::::::::::: | 1290 QQLGCGSALAA :::::::::: QQLGCGQALEA | 1350 0SL ::: TTAGTRTTSNSL 1270 |
| 1160 SCAGRLEVEYI :::::::::::::::::::::::::::::::::::: | 1230 IQCLSAPWERR.::::: | 1280 SWDLAEAEVVC ::::::::::::::::::::::::::::::::::: | CSG :::: RCSGVRTTLPT 1260 |
| 1150 ALRLYSETETE: :::::: ALRMVSEDQQ | 1210 1220 13 MWVDDIQCPKTHISIWQCLSAPW :::::::::::::::::::::::::::::::::::: | 1270 AGSWGTVCDDS .::::::::: NGSWGTVCDDS | 1330 1340 DCHAKPWGQSDCGHKEDAGVRCSG:::::::::::::::::::::::::::::::::::: |
| 0 1140 1150 DAGVICSEFTALRLYSETETE :::::::::::: DAGVICSEFLALRMVSEDQQ- | 0 1210 TGSGFMWVDDI(:: ::: | 1260 ECSGRVEIWH:::::::::::::::::::::::::::::::::::: | 1330 IDCHAKPWGQS ::::::: IDCVAEPWGQS |
| 1120 113 PSRGWGQHDCRHKE:::::::::::::::::::::::::::::::::::: | 1200 IGVVSLAPLSKT : | 1250 IRVRGGDT .:.:::: EKLRIRGGDS | 1310 1320 1330 1340 Hum. LDDMRCKGNESFLWDCHAKPWGQSDCGHKEDAGVRCSG :::::::::::::::::::::::::::::::::: |
| 1120 Hum. PSRGWG ::::: | 1190 Hum. CGENG ::::: WC1 CGDSGS | Hum WC1 CTDF 1150 | 1310 Hum. LDDN ::. WC1 LDEV |

WC1 EVPVPGTPSPSQGNEEEVPPEKEDGVRSSQTGSFLNFSREAANPGEGEESFWLLQGKKGDAGYDDVELSA WC1 LGSLLFLVLVILVTQLLRW-RAERRALSSYEDALAEAVYEELDYLLTQKEGLGSPDQMTDVPDENYDDAE --SLEENLFHEME ---DISITEA 1410 1440 -TRRRG----DTPNHGCEDAS---1400 1430 --QKHLPLRVS--1390 -LKREDPHGTRTSD-Hum. LSSIFGLLLLVLFILFLTWCRVOK---1420 1380 1380 1300 1370 Hum. TC----

Fig. 21

WC1 LGTSPVTFS

Hum. LPASEAT-K

1450

| ATGATGCTGCCTCAAAACTCGTGGCATATTTGGAAGATGCTGCTGTCATCAGAACCTTTTCTCTG:::::::::: | 80 90 100 110 120 130 140 CTGTGGTAACTTGCATCCTGCTTTCTCATCAGCAGTTTTAATGGAACAGATTTGGA : :::::::::::::::::::::::::::::::::: | 150 160 170 180 190 200 210 GTTGAGGCTGGTCAATGGAGACGGTCCCTGCTCTGGGACAGTGGAGGTGAAATTCCAGGGACAGTGGGG : ::::::::::::::::::::::::::::::: | 220 230 240 250 250 270 ACTGTGTGTGTGTGCAAACAGCTTGGATGTCCATTT :::::::::::::::::::::::::::::::: |
|--|---|---|---|
| <pre>Hum. ATGATGCTGCCTCAAAACTCGTGGCATATTGATTT ::: ::: WCl ATGGCTC-TGG 10</pre> | 80 90 100 Hum. CTGTGGTAACTTGCATCCTGCTCCTGAATTCCTGCT ::::::::::::::::::::::::::::::: | 150 160 170 Hum. GTTGAGGCTGGTCAATGGAGACGGTCCCTGCTCTG : ::::::::::::::::::::::::::::::::::: | 220 230 240 Hum. ACTGTGTGATGATGGGTGGAACACTACTGCCT- ::::::::::::::::::::::::::::::::::: |

Fig. 2Qi

| 340 | Hum. ICITICGCCAIGITICGITITIGGACAAGCCGIGACIAGACAIGGAAAAAITIGGCTIGAIGAIGTITC | •••••• | WC1 GCCATIGGITITCCIGGAGGGGCITAITITGGGCCAGGACTIGGCCCCATITGGCTTITGTATACIIC | 280 | 410 | CIGITATGGAAATGAGTCAGCTCTCTGGGAATGTCAACACCGGGAATGGGGGAAGCCATAACTGTTATCAT | ••••••••••••••••••••••••••••••••••••••• | ATGTGAAGGGACAGAGTCAACTGTCAGTGACTGTGAGCAT-TCTAATATTAAAGAC-TATC-GTAATGAT | 350 | 480 | GTGAACTGTTATGGTGAAGCCAA-TCTGGGTTTGAGGCTAG-TGGATGGAAAC | •••••• | GGCTATAATCATGGTCGGGATGCTGGAGTAGTCTGCTCAGGATTTGTGCGTCTGGCTGGAGGGATG | 420 | 550 | AACTCCTGTTCAGGGAGAGTGGAGGTGAAATTCCAAGAAAGGTGGGGGGACTATATGTGATGATGGTGGA | | WC1 GAC-CCTGCTCAGGGCGAGTAGAAGTGCATTCTGGAGAAGCTTGGATCCCAGTGT-CTGATGGGAACT | 480 |
|-----|--|---|--|-------|-------|---|---|--|-------|-------|---|--------|--|-----|-----|---|-----|--|--------|
| 330 | ATTIGGCTIG | • | ATTIGGCTTT | 270 | 400 | SAAGCCATAA | ••••••••••••••••••••••••••••••••••••••• | AAAGAC-TAT | 340 | 470 | 3GCTAG-T | •• | SCGTCTGGCT | 410 | 540 | IATATGTGAT | •• | CCAGTGT-CT | 470 |
| 320 | ACATGGAAAA | •• | ACTTGGCCCC | 260 | 390 | CGGGAATGGG | ••• | -TCTAATATT | 330 | 460 | TEGGTTTGA(| | CAGGATTTGT | 400 | 530 | GGGGGGAC | | AGCTTGGATC | 460 4. |
| 0 | STGACTAGE | ••••••••••••••••••••••••••••••••••••••• | TEGGCCAGG | 250 | 380 | ATGTCAACAC | ••• | TGTGAGCAT- | 320 | 450 | SAAGCCAA-TO | | STAGICIGCIC | 390 | 520 | FTCCAAGAAA (| ••• | CTGGAGA | 4 |
| 310 | GGACAAGCCC | • | GGGCTTATT | 240 | 370 | TCTCTGGGAA | | TGTCAGTGAC | 310 | 440 | TGTTATGGT | • | TGCTGGAC | 380 | 510 | GAGGTGAAA | ••• | GAAGTGCAT | 450 |
| 300 | STTTCGTTTT | • | TTCCTGGAG | 230 2 | 360 3 | ATGAGTCAGC | | CAGAGTCAAC | 300 3 | 430 4 | rggtgtgaac | | rggrcgggA- | 370 | 200 | AGGGAGAGTG | | AGGCCGAGTA | 440 |
| 290 | TTTCGCCAT | ••• | CATTGGT | 2 | | GTTATGGAA | •••••• | GTGAAGGGA | 30 | | GAAGATG | | CTATAATCA | 360 | 490 | CTCCTGTTC | | C-CCTGCTC | 430 |
| 280 | Hum. TC | •• | WC1 GC | 220 | 350 | Hum. CI | •• | WC1 AT | | 420 | Hum. GG | •• | WC1 GG | | | Hum. AA | •• | WC1 GA | |

| EH U | * t | E | E | ტ | | Æ | ı |
|--|---|--|--|----------------------------|-----|---|--------------------------------------|
| 620 TTCTGGAGTTGT :::::: | 0 690 TGGCACT | CICIT | 760 TAACTTG | 99 | 830 | AGAGCTG | CCCTG- 650 |
| 610 ATTTCTTCT :: | 520 680 GGAATGAGTT | :::::::::::::::::::::::::::::::::::::: | 750 ATGTCACAT | TGAAGAGTTCA | 600 | GGGGAGAGT : | G |
| 600 ATCTTCTTT :::::::: | 510 670 TTATGCCAGG | | 740 CACAATGAGG | | 810 | SGAACTAACCGCTGTATG | GGGTCTGCCC-CAGAGTG- 640 |
| 570 580 590 600 610 620 TGCCGTGGTGTGCAGGCAACTAGGATGTCCATCTTTTTTTT | 510 640 650 660 670 680 690 GTATTGCGCCCCATTTGGCTGGATGACATTTTATGCCAGGGAATGAGTTGGCACACT | :::: :: | 710 720 730 740 750 760 CATCGTGGAAATCATGACTGCAGTCACATGAGGATGTCACATTAACTTGT | :: :::::: -CAGGTCTGGGC- | 590 | ATCTTGAACTAAGGCTTGTAGGTGGAACTAACCGCTGTATGGGGAGAGTAGAGCTGA | GGGTCTG |
| 580 GTGCAGGCAA(::: CAG | 650 CCCATITGGC | ::::- AAGGCTG 540 | 720 TGGGGAAATC | | 790 | TAAGGCTTGT | TCT |
| 570 CTGCCGTGGT ::::: CTGCC | 0 640 TGTATTGCGC | | 710 ACATCGTGGA | T-GCC | 580 | GATCTTGAAC | GAGCCTGAGC 20 |
| 560 um. ACTIGAATACIGC .::: .: WC1 TCACACTIGCCAC | 490 500 630 Hum. TAATAGCCCTGCT | :.::: TTGTGGC 530 | 700 CTGGAATTGCAGA | CAGAGAGTCCAGT | 570 | TATGATAGTAGTG | TGTGAGGGGAGGAGCCTGAGCTCT- 610 620 |
| Hum. Z | Hum. | WC1 | Hum. (| WC1 | | Hum. | WC1 |
| | | | 5 m 4 4 | ٠. | | ,31 · | ~ |

'ig. 20iii

| | | | | • | | | | | | |
|-----|--|---|---|--|------|--|--|---|---|--|
| 006 | CGTATGCAA | GIICAGCAI 700 | 970 GATGTTGTA | GGTGGAGAT 760 | 1040 | GAACCGICA | ATGCCA 820 | 1110 ACTGCGACT | ••••••••••••••••••••••••••••••••••••••• | 880 |
| 068 | AGCTGATGT | CAGGTTGTTT 690 | 960 STCAGGGTCT(| : ::::: FIGAAGGGCA(750 | 1030 | AGACATTCCG | AGTCTGGCCA 810 | 1100 CAGATTTGGA | | { |
| 880 | GGGGGACCGTATGCCACCATAAGTGGAACAATGCTGCAGCTGATGTC | -GGGGGCACGTGTCACCACA-GTGGATCTGCT-CAGGTTGTTTGTTCAGCAT 660 670 670 | 950 TCATTTGCAC | :. : ::::: :::: :::: :::: :::::::::::: | 1020 | AATCTTTTCGGACTGCAGACATTCCGGAACC | rctg-gacaarggaggggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtg | 1090 CAGATGGAG | •• 6 | 870 870 |
| 870 | CATAAGTGGA | CACA-GTGGA 680 | 940 CTGGCTTGCC | | 1010 | ATCTTTTCTT | crcrerecerc 0 800 | 1080 GTGATCTGCI | | 600ATCTCC |
| 860 | GGTATGCCAC | CGTGTCAC 670 | 930 CTTCACTTCG | : :: CATGACAA-A 730 | 1000 | CCGGTAATGA | ACAATGGAGAGCGC 780 | 1070 CGATGTGTCT | | CTGTGGAGTT 850 |
| 850 | Gregegac | GGGGGCA(| 920 GGAACCGCA | .:.: AGTCCGGCT(720 | 066 | HCTCCTGCT | rcrg-gacaa | 1060 TCATCAAAA | | 840 840 |
| 840 | Hum. AAATCCAAGGAAGGTGGGGGACCGTATGCCACATAAGTGGAACAATGCTGCAGCTGATGTCGTATGCAA :::: :::: :::: :::: ::::: :::::::::: | TCCA | 910 920 930 940 950 960 970 Hum. GCAGITGGGAIGTGGAACCGCACTICACTICGCTGGCTTGCCTCATTTGCAGTCAGGGTCTGATGTTGTA | .::.:: :::::::::::::::::::::::::::: | 086 | TGGCTTGATGGTGTCTCCTGCTCCGGTAATGAATCTTTTCTTTGGGACTGCAGACATTCCGGAACCGTCA | wc1 GAACATT | 1050 1060 1070 1080 1090 1100 1110 ATTITGACTGTCTGAGATGTGTGTGTGTGTGTGTGTGTGTGTGT | | ATGITATCIGICGICAGCICGCGIGGAGITGCCAICICCACCCCGGAG-830 830 840 |
| , | Hum. | WC1 | Hum. | WC1 | | Hum. | WC1 | Hum. | į | MCT MCT |

| | 1120 | 1130 | 1,140 | 1150 | 1160 | . 1170 | 1180 |
|------|---|--------------|-----------------|------------|--|-----------|---|
| Hum. | AGCAGATGGAAGTAACAATTGTTCAGGGAGAGTAGAGGTGAGAATTCA-TGAACAGTGGTGGACAATATG | PACAATTGTTC | PAGGGAGAG | STAGAGGTGA | GAATTCA-TG? | AACAGTGGT | GGACAATATG |
| | ••••••••••••••••••••••••••••••••••••••• | ••• | •• | ••••• | | | ••• |
| WC1 | TGGTGGAAGAA | AGGTGATC | AGATCC | TAACAGCCC | GGIGATCAGATCCTAACAGCCCGATTTCACTGCTCTG- | i | GGGCTG |
| | 068 | 006 | | 910 | 920 | 930 | |
| | 1190 | 1200 | 1210 | 1220 | 1230 | 1240 | 1250 |
| Hum. | Hum. TGACCAGAACTGGAAGAATGAACAAGCCCTTGTGGTTTGTAAGCAGCTAGGATGTCCGTTCAGCGTCTTT | HAGAATGAACF | AGCCCTTG | STGGTTTGTA | AGCAGCTAGG? | ATGTCCGTT | CAGCGTCTTT |
| | ••••••••••••••••••••••••••••••••••••••• | • | •• | | | • | ••••••••••••••••••••••••••••••••••••••• |
| WC1 | - | GGAGTTGT | CCT-6 | CCT-GTGACT | GCC-CTGGGTGGTCCTGACTGTTCCCAT | rggrccrga | CTGTTCCCAT |
| | 940 | 950 | 096 | | 970 | 980 | 066 |
| | 1260 | 1270 | 1280 | 1290 | . 1300 | 1310 | 1320 |
| Hum. | Hum. GGCAG-TCGTCGTG | GCTAAACCTAG | TAATGAAG | SCTAGAGACA | CTAAACCTAGTAATGAAGCTAGAGACATTTGGATAAACAGCATATCTTGCACTGGG | CAGCATATC | TTGCACTGGG |
| | ••• | | • | | :: | | •• |
| WC1 | WC1 GGCAACACACCTCTGTGATCTGCTCAGGAAACCAGATCCAGGTGCTTCCCCAGTGCAACGA-CTCCG- | CTGTGATCTGC | TCAGGAA | ACCAGATCCA | GGTGCTTCCCC | CAGTGCAAC | GA-CTCCG |
| | 1000 | 1010 1 | 1020 | 1030 | 1040 | 1050 | 1060 |
| | 1330 | 1340 | 1350 | 1360 | 1370 | 1380 | 1390 |
| Hum. | AATGAGTCAGCTCT | TCTGGGACTGC | ACATATG | ATGGAAAAGC | CTGGGACTGCACATATGATGGAAAAGCAAAGCGAACATGCTTCCGAAGATCAGATG | rgcttccgA | AGATCAGATG |
| | • | •••••• | | •• | ••••••••••••••••••••••••••••••••••••••• | •• | • |
| WC1 | WC1TGTCTCAACCTA | PACAGGCTCTGC | 1 | DDB5 | GGCCTCAGAGGACA-GCGCCC | 222525- | CCTACTG |
| | 1070 | 1080 | | | 1090 | 1100 | |
| | | | | | | | |

Fig. 20v

| | 1400 | 00 | 1410 | | 1420 | 1430 | 1440 | | 1450 | 1460 |
|------|--------------------|-----------|--------|--------|-----------|-----------|--------------------|-----------------|---|--------|
| Hum. | CTGGAGTAATTTG | ATTTGT! | TCTGAT | AAGGCF | GATCTGG | ACCTAAGGC | TTGTCGGG | GCTCA | TTCTGATAAGGCAGATCTGGACCTAAGGCTTGTCGGGGCTCATAGCCCCTGTTATGG | TATGG |
| | • | | •• | ••• | •• | | ••• | | ••• | ••• |
| WC1 | WC1 CTCAGA | 1 1 1 1 1 | -CAG | CAGGCA | AGCTCCG- | -ccreere- | GACGGG | 25-5! | CAGCAGGCAGCTCCGCCTGGTGGACGGGG-GCGGTCCCTGCGCCGG | 95009 |
| 111 | 10 | | H | 1120 | 1130 | | 1140 | | 1150 | 1160 |
| | 1470 | 70 | 1480 | | 1490 | 1500 | 1510 | | 1520 | ((|
| Hum. | | AGGTGA | AATACC | AAGGAG | AGTGGGG | GACTGTGTG | TCATGACA | GATGG | GAGATTGGAGGTGAAATACCAAGGAGAGTGGGGGACTGTGTGTCATGACAGATGGAGCACAAGG-AATGC | AATGC |
| WC1 | GAGAGTGGAGATC | AGATCC | :::::: | AGGGCT | SESSETCOL | CACCATCTG | : :::: TGATGACG | : ::: GCTGG(| CITGACCAGGGCTCCTGGGGCACCATCTGTGATGACGGCTGGGAC-CTGGACGATGC | GATGC |
| | 1170 | 70 | 1180 | | 1190 | 1200 | 1210 | _ | 1220 | |
| 1530 | | 1540 | 15 | 1550 | 1560 | 1570 | 0 1 | 1580 | 1590 | |
| Hum. | Hum. A-GCTGTTGTGTG | GTGTGT? | AAACAA | TTGGG | TGTGGA- | AAGCCTATG | CATGTGTT | TGGTA | TAAACAATTGGGATGTGGA-AAGCCTATGCATGTGTTTGGTATGACCTATTTTAAAG | TAAAG |
| | ••• | ••• | • | •• | •• | | • | :: | ••• | •••••• |
| WC1 | CCGC-GTG | GTGTGC | AGGCAG | CTGGGC | TGTGGAG | AAGCCCTCA | -ATGCCAC | GGGGT | WC1 CCGC-GIGGIGIGCAGGCAGCIGGGCIGIGGAGAAGCCCICA-AIGCCACGGGGICIGCICACTICGGGG | 55555 |
| 1230 | 30 | 1240 | 12 | 1250 | 1260 | 1270 | - | 1280 | 1290 | |
| , , | 1600 | 1610 | - | 1620 | 1630 | 1640 | | 1650 | 1660 | - |
| Hum. | AAGCATCA | GGACCT | ATTTGG | CTGGA | GACGTTT | CTTGCATTG | GAAATGAG | TCAAA | Hum. AAGCATCAGGACCTATTTGGCTGGATGACGTTTCTTGCATTGGAAATGAGTCAAATATCTGGGACTGTGA | TGTGA |
| | ••• | | ••• | •• | • | ••••• | •• | ••• | | :: |
| WC1 | CAGGATCA | GGGCCC. | ATCTGG | TTGGAC | AACTTGA | ACTGCACAG | GAAAGGAG | TCCCA | WC1 CAGGATCAGGGCCCATCTGGTTGGACAACTTGAACTGCACAGGAAAGGAGTCCCACGTGTGGAGGTGCCC | TGCCC |
| • | 1300 | 1310 | - | 1320 | 1330 | 1340 | 0 1 | 1350 | 1360 | |

Fig. 2Qvi

| • | | | |
|--|---|---|---|
| : : 3AGTTC- .430 | 00 TCAAGGAC : : : . NCAATGGGA 1500 | CTGGACTG | 1940 GGCTCGATG ::::: GGGTGGAT- |
| :::::: | 18C SGTGTACTI ::::::: | 187 GTAGCCAG : :: :: GCAGACAG | 30 1 AAAAATTTG : : :: GCCACAGTG |
| :::::::::::::::::::::::::::::::::::::: | 1790 NAGACTGGAC .::::::: | 1860 scrgrggrgr : :: : : ccacgarcr 1550 | 1930 AGGATATGGAA :::::::: |
| . :::: GCAGGACGC 1410 | 1780 TGCTCGGG7 :::::::::::::::::::::::::::::::::: | 1850 AAGCTGCAG .:::. ATCACTGTG1 | 1920 SCTTCTA-CA(::: : : CCTTAGAGAA(|
| :::::: TGCAGACACAP 1400 | 1770 GCAGCAACCGC : AGGACCAGCAG | 1840 CTGGAACAGTA :- :-: : CATGGAAGACA | 1880 |
| 3CAGCACAAC 1390 | 1760 CTGGTGGGCG ::::::: | 1830 STGATGACGG : .:.: SCCGTAACCC | 1900 rgccargggr : : :: : |
| 1380 | 1750 3GGCTGAGG ::::::: | 1820 GCACAGTGT ::: ::: : GCAGTGTCT | 880 1890 CCCATCTTCTATCAT' |
| WC1 TTCCC | 1740 Hum. ACATGC .: | 1810 Hum. GGTGGC ::: | 1880 Hum. CCCATC WC1 TGGC |
| | | | WC1 TTCCCGGGCTGCAGACACAGCAGGACGCGGGGGTCATCTGCTCAG-AGTTC- 1370 |

| | 1950 | 1960 | 0 | 1970 | 1980 | | 1990 | 2000 | | 2010 |
|---------------|---------------------------------------|--------------|----------------|---|---------|-----------------|--|---------------|-----------------|---|
| Hum. | ATGTTTCC | TGTGAT | GGAGAT | GAGTCAG | ATCTCTG | GTCATGC | ATGITICCIGIGAGAGAAGACAGATCTCTGGTCATGCAGGAACAGTGGGTGGGGAAATAATGAC | GGGTG | GGGAA | ATAATGAC |
| | | | • | ••••••••••••••••••••••••••••••••••••••• | ••• | | ••••••••••••••••••••••••••••••••••••••• | • | •• | |
| WC1 | -AGAATCO | CAGTGTC | GGAAAA | CTGACAC | CTCTCT- | CIGGC | WC1 -AGAATCCAGTGTCGGAAAACTGACACCTCTCTCTGGCAGTGTCCTTCTGACCCTTGGAATTACAAC | CTGACCC | TTGGA | ATTACAAC |
| , | 1640 | 1650 | τ-4 | 1660 | 1670 | | 1680 | 1690 | • | 1700 |
| Hum. | 2020 TGCAGTCA | 2 ACAGTGA | .030 AGATGT | 2040 TGGAGTG | -ATCTGT | 2050 TCTGATG | 2020 2030 2040 2050 2060 2070 2080 TGCAGTCACAGTCACAGTGAAGATGTTGGAGTTCTGATG-CATCGGATATGGAGCTTGAGGCTTGTGGG | 20 ATGGAGC | 2070 GCTGAGG | 2080 |
| 5 | · · · · · · · · · · · · · · · · · · · | | | | , | | 4004004000004404040404040404040444440004404600444400464440004444000444400044440004444400044444 | | : :: :: | |
| 1 } | 1710 | 1 | 720 | 1730 1730 | | 1740 | 1750 | 17 | 1760 | |
| | 2090 | 06 | 2100 | 21 | 2110 | 2120 | 2130 | 21 | 2140 | 2150 |
| Hum. | | AGCAGGT | GTGCTG | GAAAAGT | TGAGGTG | AATGTCC | TGGAAGCAGCAGGTGTGCTGGAAAGTTGAGGTGAATGTCCAGGGTGCCGTGGGAATTCTGTGTGTAAT | TGGGAAT | TCTGT | GTGCTAAT |
| - | ••• | •• | | | • | | | •• | •• | ••• |
| WC1 | TGGAGGTC | SGTCGCT | GCTCTG | GGAGAGT | GGAGATC | CTTGACC | WC1 TGGAGGTGGTCGCTGCTCTGGGAGAGTGGAGATCCTTGACCAGGGCTCCTGGGGCACCATCTGTGATGAC | GGGGCAC | CATCT | GTGATGAC |
| 1770 | | 1780 | 1790 | 0 | 1800 | 1810 | 1820 | 0 | 1830 | |
| | 2160 | 20 | 2170 | 21 | 2180 | 2190 | 2200 | 22 | 2210 | 2220 |
| Hum. | | SAATGAA | CATTGC | TGAAGTT | GTTTGCA | GGCAACT' | GGCTGGGGAATGAACATTGCTGAAGTTGTTTGCAGGCAACTTGAATGTGGGTCTGCAATCAGGGTCTCCCA | GTCTGCA | ATCAG | GGTCTCCA |
| | • | ••• | ••• | : | | ••• | | ••• | •• | ••••••••••••••••••••••••••••••••••••••• |
| WC1 | WC1 CGCTGGGACCTGGA | ACCTGGA | CGATGC | CCGTGTG | GTGTGCA | AGCAGCT | CGATGCCCGTGTGGTGTAGCAGCTGGGCTGTGGAGAAGCCCTGGACGCCA | AGAAGC- | CCT(| GGACGCCA |
| 1840 | | 1850 | 1860 | C | 1870 | 1880 | 1890 | 0 | 1900 | 00 |
| 1 | | > |) i | • |) |)) } | | • | l |) |

Fig. 2Qviii

| | | 2230 | 2240 | 2250 | 2260 | 2270 | 2280 |
|------|---------------|---------------------|---------------------------------------|----------------------|--------------------|--|---|
| Hum. | | GCCTCATTT | CACAGAAA | GAACATTACACA | TCTTAATGTC | GAATTCTGGC | GAGA-GCCTCATTTCACAGAAAGAACATTACACATCTTAATGTCGAATTCTGGCTGCACTGGAGGGGA |
| | | (| • 6 | | | | |
| WC1 | | 1011CTTCCTTC. | TTCGGGACGG 1930 | GATCAGGGCCCA 1940 | MCTGGGCTGGAT | TGAAGTGAAC 1960 | CTGTCTTCTTCTTCGGGACGGGATCAGGGCCCATCTGGCTGG |
| 2 | 2290 47774 | 2300 | 2310 | 2320 ATGGGAGTGGAA | 2330 | 2340 3GTGTCATTTA | 2290 2300 2310 2320 2330 2340 2350 Him accementangeantenanaccaneccaneccaneccaneccaneccanec |
| WC1 | GTCCCA | AAGTATGGAG | | CTGGGGATGGCG | GCAACACAAC | .: :: .: .: .: .: .: .: .: .: .: .: .: . | WC1 GTCCCAAGTATGGAGGTGCCCTTCCTGGGGATGGCGGCAACACAACATGAATCATCAAGAAGATGCAGG |
| | 1980 | 1990 | 2000 | 2010 | 2020 | 2030 | 2040 |
| | 2360 | 2370 | 2380 | 2390 | 2400 | 2410 | 2420 |
| Hum. | | TTTGATCTGCTCAG | CCCACAGGCA | GCCCAGGCTGGT | TGGAGCTGAI | PATGCCCTGCT | CCCACAGGCCAGGCTGGTTGGAGCTGATATGCCCTGCTCTGGACGTGTTGAA |
| • | •• | | • | •• | | ••• | |
| WC1 | | TCTGCTCAG | AGTCATCTGCTCAGGATTTGTGC- 2050 2060 | GTCTGGC 2070 | TGGAGGAGAT 2080 | TGGACCCTGCT | GTCTGGCTGGAGGAGATGGACCCTGCTCAGGGCGAGTAGAA 2070 2080 2090 2100 |
| | 2430 | 2440 | 2450 | 2460 | 2470 | 2480 | 2490 |
| Hum. | GTGAA | Hum. GTGAAACATGCAGA | CACATGGCGC | TCTGTCTGTGAT | TCTGATTTCT | CICTICAIGC | CACATGGCGCTCTGTCTGATTCTGATTTCTCTCTTCATGCTGCCAATGTGCT |
| | •• | | | | • | ••• | |
| WC1 | GTGCA | TTCTGGAGA | AGCCTGGACC | CCAGTGTCTGAT | GGAAACTTCZ | ACACTCCCCAC | GTGCATTCTGGAGAGCCTGGACCCCAGTGTCTGATGGAAACTTCACACTCCCCACTGCCCAGGTCATCT |
| | 2110 | 2120 | 2130 | 2140 | 2150 | 2160 | 2170 |

Fig. 2Qix

| | 2500 | 2510 | 2520 | 2530 | 2540 | 2550 | 2560 |
|------|-----------|------------|-------------------|------------|-------------|---|--|
| Hum. | GTGCAGAGA | ATTAAATTG1 | rGGAGATGCC | ATATCTCTTT | CTGTGGGAGAT | CACTTTGGAA | GTGCAGAGAATTAAATTGTGGAGATGCCATATCTCTTTCTGTGGGAGATCACTTTGGAAAAGGG-AATGG |
| | •• | | :: | | | •••••• | ••••••••••••••••••••••••••••••••••••••• |
| WC1 | GTGCAGAGC | TGGGATG1 | IGGCAAGGCT | GTGTCT-GTC | CTGGGACACAT | GCCATTCAGA | GTGCAGAGCTGGGATGTGGCAAGGCTGTGTCT-GTCCTGGGACACATGCCATTCAGAGAGTCCGATGG |
| 2180 | | 2190 | 2200 | 2210 | 2220 | 2230 | 2240 |
| | 2570 | 2580 | 2590 | 2600 | 2610 | 2620 | 2630 |
| Hum. | TCTAACTTG | GGCCGAAAAC | STICCAGIGI | GAAGGGAGTG | AAACTCACCTT | GCATTATGCC | TCTAACTTGGGCCGAAAAGTTCCAGTGTGAAGGGAGTGAAACTCACCTTGCATTATGCCCCATTGTTCAA |
| | ••••• | •••••• | •••••• | ••• | | ••• | |
| WC1 | | GGCTGAAGAC | STTCAGGTGT | GATGGGGGGG | AGCCTGAGCTC | TGGTCCTGCC | CCAGGICIGGGCIGAAGAGIICAGGIGIGAIGGGGGGGGG |
| | 2250 | 2260 | 2270 | 2280 | 2290 | 2300 | 2310 |
| | 2640 | 2650 | 2660 | 2670 | 2680 | 2690 | 2700 |
| Hum. | CATCCGGAA | GACACTTGT | ATCCACAGCA | GAGAAGTTGG | AGTTGTCTGTT | CCCGATATAC | CATCCGGAAGACACTTGTATCCACAGAGAAGTTGGAGTTGTCTGTTCCCGATATACAGATGTCCGAC |
| | | | • | • | | ••••••••••••••••••••••••••••••••••••••• | |
| WC1 | TGTCCAGGA | GGCACATGT | CTCCACAGTG | GAGCTGCTCA | GGTTGTCTGTT | CAGTGTACAC | TGTCCAGGAGGCACATGTCTCCACAGTGGAGCTGCTCAGGTTGTCTGTTCAGTGTACACAGAAGTCCAGC |
| | 2320 | 2330 | 2340 | 2350 | 2360 | 2370 | 2380 |
| | 2710 | 2720 | 2730 | 0 2740 | 0 2750 | 2760 | 50 2770 |
| Hum. | TTGTGAATG | GCAAATCC | CAGTGTGA | CGGGCAAGTG | GAGATCAACGT | GCT-TGGACA | TTGTGAATGGCAAATCCCAGTGTGACGGCCAAGTGGAGATCAACGTGCT-TGGACACTGGGGCTCAC |
| | | • | •• | •••••• | | ••• | ••••••••••••••••••••••••••••••••••••••• |
| WC1 | | ACGGCACCTO | CTCAATGTGA | GGGCCAGGTG | GAGAT-GAAGA | TCTCTGGACG | TIATGAAAAACGGCACCTCTCAATGTGAGGGGCAGGTGGAGAT-GAAGATCTCTGGACGATGGAGAGCGC |
| | 2390 | 2400 | 2410 | 2420 | 2430 | 2440 | 2450 |

Fig. 20x

Fig. 2Qx

| Hum. 7 | | | , , , , | 0 | 0 | | |
|----------|--|--|---|---|---|---|--|
| | ATATTTGTCTGCAG | TTCCAGAGGG | GTTCCAGAGGGCAGTGCTTTGATCTGCTTAGAGGACAAACGGCTCCGCCTAGTGGAT | TCTGCTTAGAC | GACAAACGG | CTCCGCCTAG | rggat |
| | • | | • | | | | (|
| WC1 | TGCAGGCTCTGCGGC | CCTCAGAGGAG 0 2760 | gccicagaggagagticicciactgcicagacagcagciccgcciggiggac 50 2760 2770 2780 2790 2800 | ACTGCTCAGA(2780 | AGCAGGCAG(2790 | CTCCGCCTGG 2800 | rggac |
| Hum. G | 3130 3140 3150 3160 3170 3180 3190 GGGGACAGCCGCTGTGCCGGGAGAGTAGAGATCTATCACGACGGCTTCTGGGGCACCATCTGTGATGACG ::::::::::::::::::::::::::::::::::: | 3140 TGCCGGGAGAGG : :::::: CGGCGGGAGAGG | 3140 3150 3160 3170 3180 3190 GTGCCGGGAGAGATCTATCACGACGCTTCTGGGGCACCATCTGTGATGACG ::::::::::::::::::::::::::::::::::: | 3160 TCACGACGGC: : : : ::: TGACCAGGGC: 2850 | 3170 FTCTGGGGCA(:::::::: FCCTGGGGCA(| 3180 CCATCTGTGA ::::::::: CCATCTGTGA 2870 | 3190 TGACG :::: |
| Hum. WC1 | 3200 3210 3220 3230 3240 3250 3260 GCTGGGACCTGAGCGATGCCCACGTGGTGTGTGTGTGGGCTGTGGAGTGGCCTTCAATGCCACGGT .::::::::::::::::::::::::::::::::::: | 3210 CGATGCCCACGT :::::::::::::::::::::::::::::::::::: | 3210 3220 3230 3240 3250 3260 cgargcccacgrgrgrgrcaaagcrgggcrgrggagrggccrrcaargccacggr ::::::::::::::::::::::::::::::::::: | 3230 AAGCTGGGCT(:::::::: CAGCTGGGCT(| 3240 stggagtggcc ::::::::: stggagaagcc 2930 | 3250 CTTCAATGCC : :::::: CCTCAATGCC 2940 | 3260 ACGGT :::: |
| Hum. | 3270 3280 3390 3300 3310 3320 3330 3330 3320 3330 333 | 3280 GGGGAGGGGTCA :::: .:::: GGGCCAGGATCA | 3280 3320 3330 GGGGAGGGGTCAGGCCCATCTGGCTGGATGACCTGAACTGCACAGGAACGGAGTCC :::: ::::::::::::::::::::::::::::::: | 3300 GGCTGGATGA(::::::: GGCTGGACGA(| 3310 CCTGAACTGC ::::::::::: CCTGAACTGC 3000 | 3320 ACAGGAACGG ::::::::: ACAGGAAAGG | 3330 GAGTCC ::::: GAGTCC 0 |

Fig. 20xii

| | 3340 | 3350 | 3360 | 3370 | 3380 | 3390 | 3400 |
|------|---|----------------|--|--|-------------------|--------------|-------|
| Hum. | CACTIGIGGCAGIGCCCITCCCGCGGCGTGGGGCACGCACTGCAGGCACAAGGAGGACGCAGGGGTCA | CCCTTCCCGC | GGCTGGGGGC | AGCACGACTGC | AGGCACAAGG | AGGACGCAGG | GGTCA |
| WC1 | CACGTGTGGAGGTG | CCCTTCCCGG | :::::::::::::::::::::::::::::::::::::: | :::::::::::::::::::::::::::::::::::::: | AGACACAAGG | :::::::: | GGTCA |
| | 3020 303 | 30 3040 | 0 3050 | 0 3060 | 3070 | 3080 | |
| | 3410 | 3420 | 3430 | 3440 | 3450 | 3460 | 3470 |
| Hum. | TCTGCTCAGAATTCACAGCCTTGAGGCTCTACAGTGAAACTGAAACAGAGAGCTGTGCTGGGAGATTGGA | ACAGCCTTGAGGCT | GGCTCTACAG | AGTGAAACTGAAACAGAGCTGTGCTGGGAGATTGGA | CAGAGAGCIC | T.GCT.GGGAGA | TTGGA |
| WC1 | TCTGCTCAGAGTTC | CIGGCCCICA | GGATGG | GGTGAG-CGAGGACCAGCAG-TGTGCTGGGTGGCTGGA | ACCAGCAG-TG | recreeeree | CTGGA |
| | 3090 3100 | 3110 | 0 | 3120 | 3130 | 3140 | |
| | 3480 | 3490 | 3500 | 3510 | 3520 | 3530 | 3540 |
| Hum. | AGTCTTCTATAACGGGACCTGGGGCAGCGTCGGCAGGAGGAACATCACCACAGGCATAGCAGGCATTGTG | GGACCTGGGG | CAGCGTCGGC | AGGAGGAACAI | CACCACAGCC | ATAGCAGGCA | TTGTG |
| | ••••••••••••••••••••••••••••••••••••••• | •• | •••••• | | | | •• [|
| WC1 | GGTTTTCTACAACGGGACCTGGGGCAGTGTCTGCCGCAGCCCCCATGGAAGATATCACTGTGTCTCTGTGATC | GGACCTGGGG | CAGTGTCTGC | CGCAGCCCCA1 | rggaagatatc | ACTGTGTCCG | TGATC |
| 3150 | 50 3160 | 3170 | 3180 | 3190 | 3200 | 3210 | |
| - | 3550 | 3560 | 3570 | 3580 | 3590 | 3600 | |
| Hum. | TGCAGGCAGCTGGGCTGTGGGGAGAATGGAGTTGTCAGCCTCGCCCCTTTATCT-AAGACAGGCTCTG | CTGTGGGGAG | AATGGAGTTG | TCAGCCTCGC | CCTTTATC | T-AAGACAGG | CICIG |
| | ••• | •• | •••••• | | • | | |
| WC1 | TGCAGACAGCTTGGATGTGGGGACAGTGGAAGTCT-CAACACCTCTGTTGGTCTCAGGGAAGGTTCTA | ATGTGGGGAC | AGTGGAAG | TCT-CAACACC | CTCTGTTGGTC | TCAGGGAAGG | TTCTA |
| 3220 | 20 3230 | 3240 | 3250 | 3260 | . 3270 | 3280 | |

Fig. 20xii

| 3610 3620 3630 3640 3650 3660 3670 Hum. GTTTCATGTGGATGACATTCAGTGTCCTAAAACGCATATCTCCATATGGCAGTGCCTGTCTGCCCC : ::::::::::::::::::::::::::::::: | 3680 3690 3700 3710 3720 3730 3740 Hum. ATGGGAGCGAAGATCTCCAGCCCAGCAGAAGACCTGGATCACATGTGAAGATAGAATAAGAG- :::::::::::::::::::::::::::::: | 3750 3760TGC | 3770 3780 3800 3810 3820 3830 Hum. GGAGAGTGGAGCACGCAGGCTCCTGGGGCACAGTGTGTGATGACTCCTGGGACCTGGCCGAGGC :::::::::::::::::::::::::::::: |
|--|---|---------------------|---|
| 3610 Hum. GTTTCATG . : : WC1 GACCCCGG | 3680 Hum. ATGGGAGC :::::: WC1 ATGGAAAT 3360 | Hum WC1 TGTCCAAC | 3770 Hum. GGAGAGTG :::::: WC1 GGCGGGTG |

| rrggccag :::::: rrggcccr | CIGICACG:::::::::::::::::::::::::::::::: | ACAGTC : .:.: STAAGGAC 70 | 4090 ATTTTATCCA::::::CTCCCTGCCT |
|---|---|---|---|
| 3900 GCTTCGTTTGGC ::. :::::: GCAGCATTTGGC 3630 | 3970 CTATGGGACT :::::::::: CTGTGGGACT 3700 | 4040 3CTCTGG 5:::::: 3770 | 180 4090 -TAGCA-CTTATTT :::::::: CTGGCATCTTCTCC 3830 3840 |
| 3890 STGAGGGA(:: :: STGCGGIC) | 3960 AGTCATTT(::::: AGTCCTCC(3690 | 4030 CGTGAGGT(::::::: TGTGAGGT(3760 | 4080 TTTAGG : ::: CTCCCTGGG |
| 3860 3870 3880 3890 3900 AGCAGCTGGGCTCTGCTCTGGCTGCCCTGAGGGACGCTTCGTTTGGCCAG ::::::::::::::::::::::::::::::::: | 3930 3940 3950 3960 3970 CTGGTTGGATGACATGCGGTGCAAAGGAAATGAGTCATTTCTATGGGACTGTCACG ::::::::::::::::::::::::::::::::::: | 4000 4010 4020 4030 4040 CAGAGTGACTGTGGACACAAGGAAGATGCTGGGCGTGAGGTGCTCTGGACAGTC :::::::::::::::::::::::::::::::::::: | 4060 4090 TCACTGAATGCCTCCTCAGGT-CATTTAGCA-CTTATTTTATCCA ::::::::::::::::::::::::: |
| CTCTGCTCT:::::ccAGGCCCC | 0 CGGTGCAA : .:::: CAGTGCGG | 0 ACAAGGAA(::.::. ACGAGGAG(3740 | 4070 CCTC :::: AACAACCTC |
| 3870 GCTGTGGGC: ::::::: GCTGTGGC | 3940 TGACATGC :: .:: CGAGGTGC | 4010 TGTGGACA::::: TGCAAGCA | GCCT : :: . :GGACCAG2 |
| 3860 3870 AGCAGCTGGGCTGTGGCT ::::::::::::::::::::::::: | 3930 TGGTTGGAT ::::::: TGGCTGGAC 3660 | 4000 AGAGTGACT ::::::: AGAGCGACT 3730 | 4060 CACTGAATG ::::::: CACAGCAGG |
| 3840 3850 3860 3870 3880 3890 3900 Hum. GGAAGTGTGTCAGCTGTGGGCTGTGCTCTGGCTGCCCTGAGGGACGCTTCGTTTGGCCAG ::::::::::::::::::::::::::::::::: | 3910 3920 3930 3940 3950 3960 3970 Hum. GGAACTGGAACCATCTGGTTGGATGACATGCGGTGCAAAGGAAATGAGTCATTTCTATGGGACTGT ::::::::::::::::::::::::::::::::::: | 3980 3990 4000 4010 4020 4030 4040 Hum. CCAAACCCTGGGGACAGAGTGCACACAAGGAAGATGCTGGCGTGAGGTGCTCTGGACAGTC : .::::::::::::::::::::::::::::::::::: | 4050 GCTGAAAT |
| 3840 m. GGAAG :::: C1 TGAGG | 3910 m. GGAAC :::: CI GGAA? | 3980 m. CCAAZ : .:. C1 CGGAC | |
| Hum WC | 391 Hum. WC1 | Hum Hum | Hum. |
| | •• | | |

Fig. 2Qxv

| 4130 4140 -GTTTATTCTATTTCTCA : : : : : : : : : : : : : : : : : : : | 4180 CAGAGTTT :::::: AGCTGTGTATGAGGAGCT AGCTGTGTATGAGGAGCT | 4220 AGGAGAATTTATTCCATGA :::: : .: ::: AGATGACTGATGTCCCTGATGAAAT 4040 4050 | 4250 AAGAGAGGAC :.::: GAATGAGGAGGAAGTGC 0 |
|--|---|--|---|
| 4120 -CTGGTTCTG :::::::: CCTGGTCCTCGTCATCCTG | 4180 GCCCCTCAGAGTTT- :::::::::: GAAGATGCTCTTGCTGAAGCTGTGTATG 3960 3970 398 | 4210 GTTCT-CTCGAGGAGAATTTATTCCATGA- : :: :: :: :: :: :: :: :: :: :: :: :: : | CCTC :::: TCCTTCTCCCTCTCAGGGG |
| 4100 4110 4120 4130 4140 um. GTATCTTTGGGCTC-CTTCTCCTGGTTCTGTTTATTCTATTTCTCA : :::::::::::::::::::::::::::::::::: | 4150 4160 4170 CGTGGTGCCGAGTTCAGAAACAAACATCTGCCCCTCAGAGTTT :::::::::::::::::::::::::::::: | 4190 4200 4210 4220CAACCAGAAGGAGGGGTTCT-CTCGAGGAGAATTTATTCCATGA :: ::::::::::::::::::::::::::::: | 4230GATGGAGACCTGCCTCAAGAGAGAGACAC ::::::: TATGATGATGCTGAAGAAGTACCAGTGCCTGGAACTCCTTCTCCCTCTCAGGGGAATGAGGAGGAGTGC 4060 4070 4080 4090 4110 4110 |
| 4100 GTATCTT- : ::: GGGGTTCTCTGCCTTA 3850 | | | |
| Hum. WC1 | Hum. WC1 | Hum. WC1 | Hum. WC1 |

Fig. 20xvi

| 4290 | CCAA | | WC1 CCCCAGAGAAGGAGGACGGGGTGAGGTCCTCTCAGACAGGCTCTTTCCTGAACTTCTCCAGAGAGGCAGC | 4190 | 4330 | ATCGCTGTTGGGAGTT | | WC1 TAATCCTGGGGAAGGAGAAGAGGCTTCTGGCTGCTCCAGGGGAAGAAGGGGGATGCTGGGTATGATGAT | 4260 | | | | |
|------|------------------------|---|--|------|------|-----------------------------|---|---|------|------|-------------------|----|---|
| 42 | 1 | :: | CCTGAACTTCT | 4180 | 4320 | ATCGC | ••• | AAAGGGGATGC | 4250 | | | | |
| | rgacac | ••••• | AGGCTCTTT | 4170 | | | | CAGGGGAAG | 4240 | | CAAAA | • | TCTCG |
| 4280 | CTCAGA-TGACACCC | •••••• | CTCTCAGAC | 4160 | | TAGCGACAC | ••••••••••••••••••••••••••••••••••••••• | TGGCTGCTC | 4230 | 4350 | ccrcrgaagccacaaaa | • | CAGTGACTT |
| - | 1 1 1 1 1 | | GGGTGAGGT | 4150 | 4310 | AGATGC | | AGAGAGCTT | 4220 | | i | •• | GGAACATCC |
| 4270 | Hum. CCACATGGGACAAGAAC | ••••••••••••••••••••••••••••••••••••••• | AAGGAGGACG | 4140 | 4300 | CCATGGTTGTGAAGATGCTAGCGACAC | | GGGAAGGAGA | 4210 | 4340 | 9LOO | •• | WC1 GTTGAACTCAGTGCCCTGGGAACATCCCCCAGTGACTTTCTCG |
| 4260 | CCACATGG | • | CCCCAGAG | 4130 | | | ••• | TAATCCTG | 4200 | | CTT | | GTTGAACT |
| | Hum. | | WC1 | | | Hum. | | WC1 | | | Hum. | | WC1 |

45 / 96

92 112 367 152 487 172 547 32 127 132 427 52 187 72 PCCA E GAA V GTT TACC G GA ი ეც A SCA DGAT L CTG R AGG G GGT D GAT S AGT C TGT E GAG ¥ IGG A GCG M ATG CAG K AAG R CGT S AGT V GTC A GCT မှ ပြ I ATC R PCCT V GTC L K AAA T ACG E GAA D GAT S AGC K AAG L ACG D GAT CIC R AGA I ATC D SAC L IIG GAC FTC F မှ ည 999 9099 L CIG ဗ ဗ္ဗ S AGT N AAC A GCC E GAG L CTG CHG CHG A GCA TACT A GCC မှ ၁ FITC s TCG L F 3CC L Y GAC Y F TTT . I ATT PCCA C TGT F ACC I ATC PCCA L ₽ GCC ₩ TGG Q CAG ဗ ဗ္ဗ မှ ည r Grc O CAG R. AGA D GAT ၁ ညီ II C г ССG V GTC O CAG GAA T ACA CIT I ATA E GAG T ACC LCTG MATG r CIG R AGG CIC R CGA A GCT Y M ATG Y CAA က္ရ ည ဗ ဗ္ဗ N AAT STCGACCCACGCTCCGGTCTGTGGCTGAGC M ATG K AAG ი მგმ N AAC S AGC r နှင့် TCC CAG V GTG K AAG AAG H DGAT ۳. ا × F Y LCTA K AAG CAA H TACC ი მვვ CHC R AGG AAG V GTC CIT r CTT CAG × e TTT N AAT E GAA 360 9 9 9 9 TACT P CC CC F A GCC Y ဗ ဗ္ဗ SAGC N AAT V GTC G GGG C TGT G GGA L GGA

Fig. 3A

46/96

332 L027 272 847 292 907 252 787 212 667 232 727 312 967 DGAC CIC G GGG ი მმ s ICC S AGC A GCC **A** LCTG PCCC ဗ ဗ္ဗဗ K AAG A GCC N AAT ACA GAC T ACA K AAG CAG CIC V GTT FTTT R CGC င TGC A L CFG CAG V GTC M ATG E GAG H r r r L H E GAG V GTC K AAG V GTC ₩ TGG RCGT L F R AGA **₽** GCG Q CAG E GAA V GTC I ATC L A GCT W TGG F A GCT FTTC H s TCC I ATT CCC T ACC RGC 29 TACC E GAG F V 3TG GAC R I ATC FTTC L TTG S AGT CIC Y CGG T T H ¥ TGG V GTC V GTC CHC K AAG F V GTC S TCG က ၁၈၅ K AAG N AAC s TCT H LCTG N AAC V GTC T ACA A GCA A GCT O CAG H K AAG FTTC Y FTTC DGAC L O CAG G P I ATC **4** ACC. ACC F CCC N AAC K AAG S TCG R AGG L L H ភ ក្នុក N AAC EGAG L. CTG Q CAG M ATG CFC P CC CC V STT FTT PCCT K AAG A GCT TACT V GTC I ATC FTT A S ი მ C P E P A 3CC S s AGC G GGT P CCT FTTC ACA DGAC 9 9 9 9 CAG Q CAG A P C C C STCT GGC T ACC 980 Y TAT S V GTG F TTT S TCT FTTT EGAG V GTG ာ မြင့်င GAT CHC GGA

Fig. 3B

47 / 96

412 452 1387 472 432 M ATG CIG ဗ္ဗဗ္ဗ s AGC L CTT PCCC F STCT D GAC N AAC AGG DGAC CTG H 召 T ACC K AAA ი გმ AAC s AGC ස වූ TGG CGG CIG V GTG CAC **₽** CCC S AGT V GTT GTC > နှင့် CIG ი წემ GTA CIT PCCT **4** GGT > Ŋ Y TAT AAG CIG D GAT V GTG GAA V GTC GGA S ტ TACT DGAT က္ရ ည CIT A GCT CCT S C TGT L ACT S TCT T ACG ဗဗ္ဗ K AAG D GAC GAC TIC CIC Ē s TCC ი მინ CAG CAC V GTG CCT မီဗိ n HGC V GTG C TGT TGT P CCC CCC **4** F r GFC GTA > S TCA V GTG ဗ ဗ္ဗ ACA နှင့် အ L CIG s AGC TTT ACT CAA e gag ∨ GTG ი წ CAG GIG e Gag R CGA > E GAA S TCA GAG. V GTG T ACA I ATT Y TAT GCA s TCC ď K AAA ၁ဦဌ DGAT A GCA ACC GAG ი წმ V GTC E GAG S AGT M ATG CIT ACC. CAG S AGT E GAA CCT Q ဗ္ဗဗ္ဗ CTG R GGG ი წმ V GTG T ACC C TGT GAC E GAG CCA F ACA CIG သသ AAC L CIG TGG R CGG CAT Y TAT Y H 300 SCC P CCC GAC E GAG M ATG AGCT CIG R CGA IGI K AAA N AAC X AAG GTG GTG QTC S AGT ·QCAG д СС CAC

Fig. 3(

M ATG

CCC CCC

ဗဗ္ဗ

S AGT

A GCC

c TGT

A GCA

¥ ∏GG

e Gag

P CCA

N AAC

ი წნ<u>ი</u>

r CGG

E GAG

M ATG

Q CAG

K AAG

¥ TGG

s TCC 48 / 96

612 1867 712 S TCA L CTG SAGT GAA A A GCC V GTC ¥ TGG L TTG FTT CCT ი მმი ۳ ۲ F TTT Д $\mathbf{r}^{\mathbf{Y}}$ CIC ဗ္ဗဗ္ဗ A GCT GAT GGT CHC CTC CIG Y TAT N AAT S AGT နှင့် TCC CIG A GCA V GTC V GFC e Gag GHC s ICI ရ ၁၁၅ **A** GCC AGA TACT ø E GAA **န** T ACT R AGG N AAT L CTG V GTC R AAA L TTG Y A GCA ACC ACC P CCA TACT **A** GCC I ATT V GTC ₩ TGG CAG L TTG S TCC V GTC IATC S TACT C TGC P CCG D GAC A GCC F TTT Q CAG CIG s TCC Q CAG GIC V GTG CAA CAC AAG AAG GCG QCG H Y AGC GG STCT CCC CCC အ ၁၅ GGC GCC **₽** CIC GAC org ¥ TGG IATC S AGC ဂ ဦင GGT GAA H I ATC V GTG Y TAC 더 CAG GGC CGC P CCA ი მმ TGG E GAG H CHO S TCC 3 r CFC GEC V GTT Y TAC CCT CGG GG **A** Q CAG E GAG CGG A GCA G GGA CCC CCC GGA S CAG CAG CHI r CIG ATC I ATC A GCA DGAT A GCC STCA I ATC CAG S AGC CCA GTG ဗ ဗ္ဗ A GCT R AGG s TCC ပ္ပင္ပ V GTG CCT A GCA r CIG щ H I ATA S AGC N AAC TAC CHG **A**

Fig. 3D

2317 CCG V GTG A GCC D GAT E K GAG AAG AGG 1 ပ မို L CTG E V A * GAG GTA GCT TAA K E AAG GAA CCC G T GGC ACT S ဗဗ္ဗ CAG CAG CHC K AAG N N C AAC AAC TGC H ဗ္ဗ CAA

2396 2475 2633 2712 2791 2870 2949 TGACAGCAGCACAAAAAGACCACCTTTCTCCCCTGAGAGGAGCTTCTGCTACTCTGCATCACTGATGACACTCAGCAGGG TGATGCACAGCAGTCTGCCTCCCCTATGGGACTCCCTTCTACCAAGCACATGAGCTCTCTAACAGGGTGGGGGGCTACCC CCAGACCTGCTCCTACACTGATATTGAAGAACCTGGAGAGGATCCTTCAGTTCTGGCCATTCCAGGGACCCTCCAGAAA CACAGTGTTTCAAGAGATCCTAAAAAACCTGCCTGTCCCAGGACCCTATGGTAATGAACACCCAAACATCTAAACATC ATATGCTAACATGCCACTCCTGGAAACTCCACTCTGAAGCTGCCGCTTTGGACACCCAACACTCCCTTCTCCCAGGGTCA TGCAGGGATCTGCTCCCTCCTGCTTCCCTTACCAGTCGTGCACCGCTGACTCCCAGGAAGTCTTTCCTGAAGTCTGACC **ACCTITCTICTICCTICAGITGGGGCAGACTCTGATCCCTTCTGCCTGGCAGAATGGCAGGGGTAATCTGAGCCTTCT** TCACTCCTTTACCCTAGCTGACCCCTTCACCTCCCCCTCTCTTTTCCTTTTTGGGATTCAGAAAACTGCTTGTC

Fig. 3F

| 70 | OTLLLS STLLLS OTLLLS | 140 NVTHLY | ::: NATHLY 140 | 210 EPILMR :::::: EPILMR | 280 GEKLLQ :::::: GEKLLQ 280 |
|----|--|--|--|---|---|
| 9 | FFHQKGLQDFI :::::::: FFQQKGLRDFI | 50 130 FNFIRVLVSYN | FNFIRVLVSYN 130 | 190 200 210 AVLVDGMLYSGTMNNFLGSEPILMR :::::::::::::::::::::::::::::::::::: | 270 .VARVCKNDVG(::.::::::: .VAQVCKNDVG(270 |
| 50 | IGLFLFQLLQLLLPTTTAGGGGQGPMPRVRYYAGDERRALSFFHQKGLQDFDTLLLS : .:.::::::::::::::::::::::::::::::::: | 20 30 40 50 60 70 90 100 110 120 130 140 ILALDIQDPGVPRLKNMIPWPASDRKKSECAFKKKSNETQCFNFIRVLVSYNVTHLY | .::::::::::::::::::::::::::::::::::::: | 150 180 190 200 210 210 200 210 210 200 210 210 21 | 230 240 250 260 270 280 LRWLHHDASFVAAIPSTQVVYFFFEETASEFDFFERLHTSRVARVCKNDVGGEKLLC ::::::::::::::::::::::::::::::::::: |
| 40 | GGQGPMPRVR: | 40 110 PASDRKKSEC | :::::: PASERKKTECI 110 | 180 GOSPEDPAHK) ::::. GOSPLTLFTS' | 250 YFFFETASE :::::::::: YFFFETASE |
| 30 | LLPTTTAGG PSLPPASGT | 30 100 PRLKNMIPW | ::::::: PRLKNMIPW 100 | 170 SEDKVMEGK :::::: LIDKVMDGK | 240 AAIPSTQVV :::::::::: AAIPSTQVV 240 |
| 20 | GLFLFQLLQL .:.::.: RVFFFQLFLL | 20 90 LALDIQDPGV | ::::::: TLALNIQNPGI 90 | .160 170 170 1 IELQDSYLLPISEDKVMEGKGOSPF :::::::::::::::::::::::::::::::::::: | 230 LRWLHHDASFV :::::::::::::::::::::::::::::::::::: |
| 10 | MALPALGLDPWSL:::::::::::::::::::::::::::::::::::: | 10 80 GDGNTLYVGAREAI | DDGNTLYVGARETV 80 | C : E : E : E : E : E : E : E : E : E : | 220 240 250 260 270 280 280 280 280 270 280 280 270 280 280 270 280 270 280 270 280 270 280 270 280 270 280 270 280 270 280 270 280 220 230 240 250 250 260 270 280 |
| | Hum. Mur. | Hum. | Mur. | Hum. Mur. | Hum. Mur. |

Fig. 31

| 350 | DIERVE | DIERVE | 350 | 420 | YTRLAV | .:::::: | • • • • • • • • • • • • • • • • • • • | 420 | 490 | SESGGVW | ••• | SESGGIW | 490 | 560 | SIRPOS | ••• | RSPRROS | |
|-----|--|--|---|-------|--|--|---|-----|-----|--|-----|--|-----|-----|--|-----|--|-----|
| 340 | SAVCAFSLI | SAVCAFSL | 340 | 410 | PLLVKSGV | TOT TWEET | | 410 | 480 | APTOGAVEVO | • | NEAQGAVFA | 480 | 550 | VACASGPMSI | | VCTRGPMAI | 000 |
| 330 | TSQWQVGGTRS | | 330 | 400 | HFLMDEQVVGT | TO THE PROPERTY OF THE PROPERTY AND DESCRIPTION OF THE PROPERTY OF THE PROPERT | יס א משמשת זע | 400 | 470 | DPEPVRNLQL? | | DSEPVRNLQLA | 470 | 540 | KQDMERGNPEV | | KQDMERGNPEV | 54O |
| 320 | SPTAPHIYAVE | RUAVIARUADA | 320 | . 068 | SSDKALTFMKD | | AND STREET | 390 | 460 | AHLVEEIQLFP | •• | AYLVEEIQLSP | 460 | 530 | CLLSAPNLNSW | • | SLLSGST-KPW | 530 |
| 310 | IRHAVLLPAD | O A G T. TV A H G T | 310 | 380 | PRPGSCSVGP | | ドストらいしいいられてい | 380 | 450 | HKAVVSGDSS | | HKAVVPQDSS | 450 | 520 | CAWDPESRIC | ••• | CAWDPESRIC | 520 |
| 300 | TOPGOLPENV | | 300 300 | 370 | WITYRGPEIN | | W.T.T.Y.Cの巨くの | 370 | 440 | MYLGTTTGSL | | MYLGTSTGPL | 440 | 510 | VDCVLARDPH | | VDCVLARDPH | 210 |
| 290 | KKWTTFLKAQLLCTQPGQLPFNVIRHAVLLPADSPTAPHIYAVFTSQWQVGGTRSSAVCAFSLLDIERVF | SECTION OF THE PROPERTY OF THE PROPERTY OF THE SOUND OF THE SAUCA PSITDIERVE | 700 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 | 360 | KGKYKELNKETSRWTTYRGPETNPRPGSCSVGPSSDKALTFMKDHFLMDEQVVGTPLLVKSGVEYTRLAV | | KGKIKELNKETOKWII IKGOE VOFRFGOCOMGFOODINALI EMKODE HEDEMV VGIE DIVINGO DI LINIO | 360 | 430 | ETAQGLDGHSHLVMYLGTTTGSLHKAVVSGDSSAHLVEEIQLFPDPEPVRNLQLAPTQGAVFVGFSGGVW | | ESARGIDGSSHVVMYLGTSTGPLHKAVVPQDSSAYLVEEIQLSPDSEPVRNLQLAPAQGAVFAGFSGGIW | 430 | 200 | RVPRANCSVYESCVDCVLARDPHCAWDPESRTCCLLSAPNLNSWKQDMERGNPEWACASGPMSRSLRPQS | | RVPRANCSVYESCVDCVLARDPHCAWDPESRLCSLLSGST-KPWKQDMERGNPEWVCTRGPMARSPRRQS | 200 |
| | Hum. | N. | Eur. | | Hum. | ; | Mur. | | | Hum. | | Mur. | | | Hum. | | Mur. | |

Fig. 3(

RPQIIKEVLAVPNSILELPCPHLSALASYYWSHGPAAVPEASSTVYNGSLLLIVQDGVGGLYQCWATENG Mur. PPQLIKEVLTVPNSILELRCPHLSALASYHWSHGRAKISEASATVYNGSLLLLPQDGVGGLYQCVATENG FSYPVISYWVDSQDQTLALDPELAGIPREHVKVPLTRVSGGAALAAQQSYWPHFVTVTVLFALVLSGALI Mur. YSYPVVSYWVDSQDQPLALDPELAGVPRERVQVPLTRVGGGASMAAQRSYWPHFLIVTVLLAIVLLGVLT Mur. Illasplgalrargkvogcgmlpprekaplsrdohlopskdhrtsasdvdadnnhlgaeva ILVASPIRALRARGKVQGCETLRPGEKAPLSREQHLQSPKECRTSASDVDADNNCLGTEVA 069 760 680 750 610 670 900 740 670 099 590 730 99 580 720 580 650 640 710 700 Hum. Hum. Hum.

Fig. 3H

Fig. 31

| 250 260 310 ATGGAAATACTCTACGTGGGGGCTCGAGAAGCCATTCTGGCCTTGGATATCCAGGATCCAGGGTCCC :::::::::::::::::::::::::::::::: | gaag :::: gaag | 390 440 450 AGCAATGAGACACAGTTTCATCCGTGTTTCTTACAATGTCACCCATCTTACACT ::::::::::::::::::::::: | 460 470 480 490 500 510 520 GCGGCACCTTCGCCTTGTACCTTCATTGAACTTCAAGATTCCTACCTGTTGCCCATCTC : ::::::::::::::::::::::::::::::::: |
|---|--|---|--|
| 310 GGGG :: | 380 Agaa : : : : Agaa | 450 CTAC ::: | 520 CCCA :::: CCCA |
| ::: ::: cccA | TTTA:::::::::::::::::::::::::::::::::: | ATCT:::ACCT | GTTG :::: GTTG 550 |
| 0 AGGA :::: AGAA | 0 HGCC HGCC | O ACCC :::: | ACCT |
| 300 ATCCA: ::::: ATCCA: | 370 AATGT(::::: AATGT(0 | 440 TGTCA(:: : TGCTA(| 510 TCCTAC |
| GGATAS GAATAS 330 | AGTGAL : :: ACCGAL 400 | ACAATC | AGATT(::::AGATT() |
| 290 CCTT :::: | 360 AAAG :::: | 430 TCTT:::::::::::::::::::::::::::::::::: | 500 TTCP : :: |
| CTGG :::: :::: 320 | AGAAA ::::: AGAAA 390 | GGTT : : : GGTC 460 | rgaac ::::: rgaac 530 |
| 250 260 310 ATGGAAATACTCTACGTGGGGGCTCGAGAAGCCATTCTGGCCTTGGATATCCAGGATCCAGGGGTCCC ::::::::::::::::::::::::::::::: | 320 330 340 350 360 370 380 AGGCTAAAGAACATGATACCGTGGCCAGCCAGTGACAGAAAAAAGAGTGAATGTGCCTTTAAGAAGAAGAG ::::::::::::::::::::::::: | 390 400 450 450 430 440 450 agcaatgacaccatctacacctacctacctaccatctacacctacctaccatctacaccatctacaccatctacaccatctacaccatctacaccatctacaccatctacaccatctacaccatcaccatctacaccatcacaccatcacaccatcacacaccac | 0 480 520 TTCAGCCCTGCTTGTACCTTCATTGAACTTCCAAGATTCCTACCTGTTGCCCATCT ::::::::::::::::::::::::::::::::::: |
| 28 AGAAG(::::. AGAGA(310 | 3(GCCAG' ::::: GCCAG' 380 | 4; TCCGT(::::. TTCGA(| TACCI |
| TCGA :::: TCGA | CCAG :::: CCAG | TCAT TCAT | TT GT: |
| 270 9660 1::: | 340 GTGG ::: CTGG | 410 AACT :::: AACT | 480 CTGC CTGC |
| GTGG :::: GTGG | TACCO | TTTC2 | AGCC. 510 |
| 250 260 ATGGAAATACTCTTAC :::::::::: ATGGCAACACTCTCTAT 0 | 330 CATGA ::::: | 400 CAGTG ::::: CAGTG | 470 CCTTC ::::: |
| 2 ACTCT ::::: ACTCT 290 | 3 AGAAC ::::: AGAAC 360 | 430 | 4 TTCGC :: :: TTTGC 500 |
| . : : : 2ACP | TAAA :::: TAAA | .TGAG | 460 GGCACCIT :::::: GGGACCIT |
| 250 NTGGA | 320 346601 3 : : : : : : : A6601 | 390 GCAA::::: | 460 3CGGC/ 1: 1TGGG/ |
| Hum. A. : Mur. A. 280 | 320 330 340 350 360 370 380 Hum. CAGGCTAAAGAACATGATACCGTGGCCAGCCAGTGAAAAAAAA | 390 400 410 420 430 440 450 Hum. AGCAATGACATCATCCGTGTCCTGGTTTCTTACAATGTCACCCATCTACACCT ::::::::::::::::::::::: | Hum. G : Mur. G' 490 |
| Hr. Mu | H Z | H M | Hr. Mu |

| 810 820 830 840 850 850 870 870 850 860 870 870 | AGAGCTGTATATATCCAGGGTGGCTCAAGTCTGCAAGAACGACGTGGGCGGTGAAAAGCTGCTGCAGAAG 880 850 850 900 | 880 890 910 910 920 940 Hum. AAGTGGACCACCTTCCTGAAGGCCCAGCTGCTCTGCACCCAGCCGGGGCAGCTGCCCTTCAACGTCATCC ::::::::::::::::::::::::::::::::: | 950 960 970 980 990 1000 1010 GCCACGCGGTCCTGCTCCCGATTCTCCCACAGCTCCCCAGTG ::::::::::::::::::::::::::::::::::: | 1020 1030 1040 1050 1060 1070 1080 GCAGGTTGGCGGACCTCTGGGGGTTTGTGCCTTCTCTCTTTGGACATTGAACGTGTCTTTAAG ::::::::::::::::::::::::::::::::: |
|---|--|--|--|--|
| 860 CGGCGAAAAG | ::: :::: CGGTGAAAAG 890 | 930 CAGCTGCCCT :::::::: CAGCTGCCAT | 1000 ACGCAGTCTT :::::::::::::::::::::::::::::::::: | 1070 GGACATTGAA :::::::::: GGACATTGAG |
| 850 ATGACGTGGG | : :::::: ACGACGTGGG 880 | 920 CCAGCCGGGG ::::::::: TCAGCCAGGG | 990 CCCCACATCT :::::::: | 1060 TCTCTCTCTT :::::::: TCTCTCTCAC |
| 840 GTCTGCAAGA | :::::::::::::::::::::::::::::::::::::: | 910 TGCTCTGCAC ::::::::: TGCTCTGCGC | 980 TCCCACAGCT ::::::::: TCCCTCTGTT 1010 | 1050 GTTTGTGCCT :: :::::: GTCTGTGCCT 1080 |
| 830 GGTGGCTAGA | GGTGGCTCAA 860 | 900 AAGGCCCAGC ::::::: AAAGCCCAGT | 970 CCGCCGATIC ::::::::: CCGCCGATIC | 1040 ::::::::::::::::::::::::::::::::::: |
| 820 ACACATCGCG | ATATATCCAG 850 | 890 CACCTTCCTG ::::::: CACCTTCCTC 920 | 960 GTCCTGCTCC :::::::: GTCCTGCTGC | 1030 GCGGGACCAG :::::::: GCGGGACCAG |
| | 00 | 880 AAGTGGAC(::::::: AAGTGGAC(910 | O1 | 1020 10. . GCAGGTTGGCGGGA ::::::::::::: . GCAGGTTGGCGGGA |
| ä | Mur | Hum. Mur. | Hum. Mur. | Hum. Mur. |

| | | | | | | | | | | | | | , |
|------|--|--|---|---|--------------|---|---|------|------|--|---|--|------|
| 1150 | GGGAAATACAAAGACTTGAACAAAGTTCACGCTGGACTACTTATAGGGGCCCTGAGACCAACCCCC | GGGAAGTACAAGAACTACCGCTCCCGCTGGACCACTTACCGGGGCTCAGAGGTCAGCCCGA 1130 1140 1150 1160 1160 | 1160 1170 1180 1190 1200 1210 1220 Hum, GGCCAGGCAGTTGCTCAGCCCCTCCTGATAAGGCCCTGACCTTCATGAAGGACCATTTCCTGAT | GGCCAGGCAGTTGCTCCATGGGCCCCTCTGACAAAGCCTTGACCTTCATGAAGGACCATTTTCTGAT | 1250 1290 | GGATGAGCAAGTGGTGGGACGCCCCTGCTGGTGAAATCTGGCGTGGAGTATACACGGCTTGCAGTGGAG | Mur. GGATGAGCACGTGGTAGGAACACCCCTGCTGGTGAAGTCTGGTGTGGAGTACACACGGCTTGCTGTGGAG | 1320 | 1360 | ACAGCCCAGGGCCTTGATGGGCACAGCCATCTTGTCATGTACCTGGGAACCACCACAGGGTCGCTCCACA | | TCAGCTCGGGGCCTTGATGGGAGCAGCCATGTGGTCATGTATCTGGGTACCTCCACGGGTCCCCTGCACA | 1390 |
| 1140 | GAGITGAACAAAGITCACGCIGGACIACITAIAGGGGCCCIGAGACCAACCC | CCGGGCTCAG | 1210 FTCATGAAGGA | SCTCCATGGGCCCCTCTTGACAAAGCCTTGACCTTCATGAAGGACCATTTT | 1240 1280 | AGTATACACGG | AGTACACACGG | 1310 | 1350 | AACCACCACAG | ••••••••••••••••••••••••••••••••••••••• | PACCICCACGG | 1380 |
| 1130 | GGACTACTTA | GGACCACTIA 1160 | 1200 GGCCCTGACC | AGCCTTGACC | 1230 1270 | TCTGGCGTGG | TCTGGTGTGG | 1300 | 1340 | TGTACCTGGG1 | | TGTATCTGGG | 1370 |
| 1120 | ACTTCACGCT | ACCTCCCGCT 1150 | 1190 CCTCTGATAA | CCTCTGACAA | 1220 1260 | GCTGGTGAAA | GCTGGTGAAG | 1290 | 1330 | CATCTTGTCA | ••• | CATGTGGTCA | 1360 |
| 1110 | GAACAAAGAA | GAACAAGGAG 1140 | 1180 Gregeccerr | ATGGGCCCCT | 1210 1250 | GGACGCCCCT | GAACACCCCT | 1280 | 1320 | TGGGCACAGC | ••• | TGGGAGCAGC | 1350 |
| 1100 | ACAAAGAGTT | ACAAGGAGCT | 1170 CAGTIGCICA | CAGITGCICC | 1200 1240 | CAAGTGGTGG | CACGTGGTAG | 1270 | 1310 | AGGCCCTTGA | •• | GGGCCTTGA | 1340 |
| | | - | 1160 m. GGCCAGG | • | 1190 | Hum. GGATGAGCAAGT | r. GGATGAG | 1260 | 1300 | Hum. ACAGCCC | ••• | | 1330 |
| | HUH. | Mur. | Hui | Mur. | | Hui | Mu | | | Hui | | Mur. | |

| 1370 1380 1390 1400 1410 1420 1430 AGCTGTAGCTGTACCCTGACCTGAACC | AGGCTGTGGTGCCTCAGGACAGCAGTGCTTATCTCGTGGAGGAGATTCAGCTGAGCCCTGACTCTGAGCC | 00 1410 1420 1430 1440 1450 1460 1440 1450 1460 1470 1480 1490 1500 | :::::::::::::::::::::::::::::::::::::: | 1510 1520 1530 1540 1550 1560 1570 GTGCCCGAGCCAACTGTCTATGAGAGCTGTGTGGACTGTGTCCTTGCCCGGGACCCCCACTGTG :::::::::::::::::::::::::::::::::: | 1580 1590 1600 1610 1620 1630 1640 CCTGGGACCCTGAGTCCCGAACCTGTTGCCTCCTGCCCCCAACCTGAACTCCTGGAAGCAGGACAT ::::::::::::::::::::::::::::::::::: |
|--|--|--|--|--|--|
| 1420 PAGCTGTTCCC | AGCTGAGCCC | 1450 1490 3CTTCTCAGGA | :::::::::::::::::::::::::::::::::::::: | 1560 :::::::: 3CTTGCCAGGG 1590 | 1630 CTGAACTCCTG ::::: CAAGCCTTG |
| 1410 GGAAGAGATIC | :::::::::::::::::::::::::::::::::::::: | 1440 1480 STGTTTGTAGG | :::::::::::::::::::::::::::::::::::::: | 1550 IGGACTGTGTC::::: IGGACTGTGTGT | 1620 TGCCCCCAACC :: : :::: TGGCTC-TACC |
| 1400 CTCATCTGGT | :: :::: :: CITAICICGIO | 1430 1470 CCAGGGTGCA | :::::::::::::::::::::::::::::::::::::: | 1540 GAGAGCTGTG' :::::::: GAGAGCTGTG' 1570 | 1610 GCCTCCTGTC::::::::::::::::::::::::::::: |
| 1390 GACAGCAGTG | GACAGCAGTG | 1420 1460 TGGCCCCCAC | :::::::::::::::::::::::::::::::::::::: | 1530 TAGTGTCTAT ::::::: CAGTGTCTAC | 1600 CGAACCTGTT::::::. |
| 1380 GGTAAGTGGG | GGTGCCTCAG | 1410 1450 AACCTGCAGO | GTTCGAAACCTGCAGC | 1520 GAGCCAACTG ::::::::: GGGCCAATTG | 1590 .cccrgagrcc :::::::: .cccrgaarca |
| 1370 Hum. AGGCTGT | | 1400 1440 Him TGTTGG | 4 | 1510 1: Hum. GTGCCCGAGCCA/ :::::::::: Mur. GTTCCCAGGGCCA/ 1540 1550 | 1580 1. Hum. CCTGGGACCCTGA :::::::::::::::::::::::::::::::::::: |
| ,11, | . 4 | 1 | - 4 | щ Д | |

Fig. 3N

| 1650 GAGCGGG | 1660 GGAACCCAG | 1650 1660 1670 1680 1690 1700 1710 1650 1700 1710 1710 1710 1690 1710 1710 1710 | 1680 SCCAGTGGCCC | 1690 CATGAGCAGG | 1700 AGCCTTCGGCC | 1710 CTCAGAGCCGC |
|---|--|---|---|---|--|---|
| B | :::::::::::::::::::::::::::::::::::::: | :::::::::::::::::::::::::::::::::::::: | :::::::: \cccereccc 1710 | :::::::::::::::::::::::::::::::::::::: | | stcagageeee 1740 |
| 1720 GCAAAT :::: TCAACT 750 | 1720 1730 CCGCAAATCATTAAAGA :::::::::::: CCTCAACTAATTAAAGA 1750 1760 | 1720 1730 1740 1750 1760 1770 1780 CCGCAAATCATTAAAGAAGTCCTGGCTGTCCCAACTCCTGGAGCTCCCCTGCCCCCACCTGTCAG :::::::::::::::::::::::::::::::::::: | 1750 rcccaacrcc ::::::::: rcccaacrcc 1780 | 1760 ATCCTGGAGC ::::::::: ATCCTGGAGC 1790 | 1770 ::::::::::::::::::::::::::::::::::: | 1780 CCACCTGTCAG ::::::::: CCACCTGTCAG 1810 |
| 1790 CTTGGCC ::::: ACTGGCC | 790 1800 TGGCCTCTTATTAT :::::::::: TGGCCTCTTACCAC | 1790 1840 1850 1850 1850 1850 1840 1850 CCTTGGCCTCTTACACTGTCTACAATGGCCTTTGGCCTCTTTCCACTGTCTACAATGGCCTTGGCCTCTTTCCACTGTCTACAATGGCCTTGGCCTCTTTTACACTGTCTACAATGGCCTGGCCTCTTACAATGGCCTGGAGCCTCTACAAATCTCAGAAGCCTCTGCTACCGTCTACAATGGCCAAAATCTCAGAAGCCTCTGCTACCGTCTACAATGGCTACCAATGGCTACCAATGGCCTGGAGCTACCAATGGCTACCAATGGCTACCAATGGCTACCAATGGCTACCAATGGCTACCAATGGCTACCAATGGCTACCAATGGCTACCAATGGCTACAATGGCTACCAATGGCTACCAATGGCTACCAATGGCTACCAATGGCTACCAATGGCTACCAATGGCTACAATGGCTACCAATGGCTACAATGAATG | 1820 cccagcagcag ::::::: | 1830 TCCCAGAAGC :::::::::::::::::::::::::::::: | 1840 :::::::::::::::::::::::::::::::::::: | 1850 STCTACAATGG :::::::::: STCTACAATGG 1880 |
| 1860 TCCCTCT ::::::: TCCCTCT | 1860 1870 CTCCCTCTTGCTGATAG ::::::::::::::::::::::::::::::::: | 1860 1870 1880 1890 1900 1910 1920 CTCCCTCTTGCTGATAGTGCAAGTTGGGGGTCTCTACCAGTGCTGGGCAACTGAGAATGGCTTT :::::::::::::::::::::::::::::::: | 1890 STTGGGGGTCT :: :::: :: STCGGGGGCCT | 1900 CTACCAGTGC ::::::::: CTACCAGTG1 | 1910 TGGGCAACTG ::::::: GTGGCGACTG | 1910 1920 GGCAACTGAGAATGGCTTT ::::::::::: GGCGACTGAGAACGGCTAC 1940 1950 |

| TCATACCCTGTGGTCTCCTATTGGGTAGACAGCCCAGGCCCCTGGCGCCTGGACCCTGAGCTGGCGG 1960 1970 1980 2000 2010 2020 2000 2010 2020 2030 2040 2050 2060 GCATCCCCGGGGAGCATGTGAAGGTCCCGTTGACCAGGGTCAGTGGGGGCCGCCCTGGCTGCCCAGCA |
|---|
| ::.: ::::: ::::::::::::::::::::::::::: |
| 2070 |
| GICCTACTGGCCCCCACTTTGTCACTGTCACTGTCCTTTTGCCTTTAGTGCTTTCAGGAGCCCTCATCATC :::::::::::::::::::::::::::: |
| GTCCTACTGGCCCCATTTTCTCATCGTTACCGTCCTGGCCATCGTGCTCCTGGGAGTGCTCACTCTC 2100 2110 2120 2130 2140 2150 2160 |
| 2140 2150 2160 2170 2180 2190 2200 CTCGTGGCCTCCCCATTGAGAGCTCCGGGCTAGGGTTCAGGGCTGTGAGACCCTGCGCCCTG |
| |
| CTCCTCGCTTCCCCACTGGGGGCGCTGCGGGCTTCGGGGTTCAGGGCTGTGGGATGCTGCCCCCCAA 2170 2180 2190 2190 2200 2200 |

Fig. 3F

| ### GGAGAAAGGCCCGTTAAGCAAGACACCTCCCAAGGAATGCAGGACTCTGCCAGTGA ################################# | 2260 2270 CAGGACCTCTGCCAGTGA | AGGACCTCTGCCAGTGP 90 2300 | 2330 2340 ACAGG-CCGGGGCTGC | ACAGATCCGCAGCTGAGC 2360 2370 | 2400 2410 SATGACAGCACAAAA .:: .:.:. 2420 | 2470 2480 ACTCAGCAGGGTGATGC :: :.::: AC-CTACAGC |
|---|---------------------------------|--|-------------------------------|---------------------------------|---|--|
| 2210 2 GGGAGAGGCCCC :::::::::::::::::::::::::::::::: | 2250 22 CTCCCAAGGAATGCF | | 2320 23 TTAAACTCTAGGCAC | TTAAACA-GGGACAC 2350 | 2390 ACAGCCTGACTAGC :::: ACAG | 2460 GCATCACTGATGACI :::: GCTA-ACGTGTGTCZ |
| 2210 2 GGGAGAGGCCCC :::::::::::::::::::::::::::::::: | 10 2240 GCAACACCTCCAGT | CCAGCACCTCCAGC | 10 2310 GGCACTGAGGTAGC | | 2370 2380 TGGGCGGCCCAAGC : . : : :::: TATGCCAGGC | 2440 2450 |
| | 2220 223 | :::::::::::::::::::::::::::::::::::::: | 2290 230 | ACAACAACCATCTG 2320 2330 | 2360 CCTGGCCATGCTGGC :::::::::: | 2430 rcrcccrgagagga : : : : : |
| | | | | CGTAGAT 2310 | | |

Fig. 3C

| 0 | E! • | · | ଓ ∙ ୟ | لا ، ۵ | 0 द : स |
|------|---|--|--|--|---|
| 2550 | | - 666CT | ATTCCAG ::: GACCCAA | 2670 2680 TGCCTGTCCCAGGACCCTATGGTA :::::::::::::::::::::::::::::::: | 2710 2720 2730 2740 2750 1CTBAACAATCATATGCTAA-CATGCCACTCCTGGAAACT-CCACTCTGAA ::::::::::::::::::::::::::::::::::: |
| | GIGG | | CCAT TTGA 2580 | 2680 ACCCIA : ::. | CACT |
| 2540 | CAGG | | 2610 CAGTTCTGGC ::::: TGATTTTGAT | 26 GGAC ::: | 2740 MAACT-C :: |
| 7 | CTAA | | TCAGTT::: | CCCA: : . CT-G | 274 GAAAC : |
| 0 | CICI | | 00 CTTC ::: CTTT | 2670 CTGT :::: CTGT | CCTG |
| 2530 | TGAG | | 2600 AGGATCCTT ::::::::: AGGTTTCTT 60 | CTGC :: GTGG | 2730 ACTC .::: |
| ٠ | CACA | :::::::::::::::::::::::::::::::::::::: | AGAG :: C-AG | 2640 2650 2660 ACACA-GTGTTTCAAGAGATCCTAAAAAAACC: :::::::::::::::::::::::: | CA |
| 2520 | CAAG | ::: -aag | 2570 2580 2590 TCCTACACTGATA-TTGAAGAACCTGGA .: : .: :: :: :: :: :: :: :: :: :: :: :: | 26 AAAA : TGAA | TGC- |
| 7 | CTAC | ::::: TTCTGC 2500 | GAAC :::: GAGC 0 | 0 :ATCCT : ::: :A-CCA 2620 | A-CAT |
| 0 | CCTT | TCTT 25 | 2580 -TTGAAG :::: CAGGAAG | 2650 GAGA' :::: GAGA' | 2720 GCTA :::: |
| 2510 | ACTC | ACTC | 25 TA-T | TCAA | CATAT |
| | TGGG | :::: TGGG 2490 | ACTGA .:::. GCTGT 2540 | 2640 -GTGTT :::: CGTGCT 2610 | 2710 ACAATC .:::: |
| 2500 | CCTA | CCTG | 2570 CTACA : TTGTG | 26 CA-G : : | 27 AAAC .:.: GAGC |
| 7 | CICC | :::: TCCTCC 2480 | CICC CICC .: TACT | 0 .GAAACA : . : : : : 'GTAACA | ATCTA :::: |
| | TG-C | GGTC 24 | 0 CCTGC :::: CCTGT | 2630 CCAGA ::::: | 90 2730 2740 2750 IGAACACCAAACATCTAAACAATCATATGCTAA-CATGCCACTCCTGGAAACT-CCACTCTGAA .:::::::::::::::::::::::::::::::::: |
| 2490 | AGIC | :::: CAGTAG 0 | 2560 CCCCAGACCTG : :::::: CTCCATACCTG | CT-C ::: | ACACC ::: AC-CT |
| | ACAGCAGTCTG-CCTCCCCTATGGGACTCCCTTCTACCAAGCACATGAGCTCTCTAACAGGGTGGGGGGCT | :: :::: :::::::::::::::::::::::::::::: | 2560 2570 2580 2590 2600 2610 ACCCCCAGACCTGCTCCTACACTGATA-TTGAAGAACCTGGAGAGGATCCTTCAGTTCTGGCCATTCCAG .:::::::::::::::::::::::::::::::::::: | 2620 2630 2640 2650 2660 2670 2680 Hum. GGACCCT-CCAGAAAGACA-GTGTTTCAAGATCCTAAAAAACCTGCCTGTCCCAGGACCCTATGGTA ::::::::::::::::::::::::::::::::::: | 2690 2740 2720 2730 2740 2750 Hum. AIGAACACCAAACATCAAATCATAIGCIAA-CAIGCCACICCIGGAAACI-CCACICIGAA : .::::::::::::::::::::::::::::::::: |
| | Hum. A | | | 2620 ii. Gi | 26 Hum. A : : Mur. A |
| | Huī | Mur. | Hum. Mur. | 26 Hum. Mur. | Hu |

Fig. 35

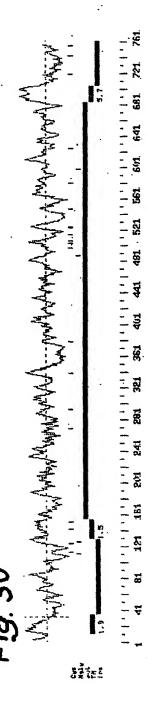
4

Hum. AAAAAAAGGGCGGCCGC

Mur. GCTTTAAAG-

3100

3090



86 392 66 332 106 452 126 512 **₽** N AAT GGG CAG င TGC CIT Y C TGT ტ GG. FTTC M ATG PCCA PCCG မှ င် H င် မြင့် CAG GAC TGG င္ပင္ပင္က K AAG ස පු G GGA S TCA ය ප K AAA E GAG F EC R AGG CAG G GGA r Agg Y Y မှ ည A GCC R CGG R AGG PCCA M ATG s Icc ¥ FGG I ATC T ACC DGAC P CCA E GAA SCGAAGCGCCCTGCGACCCGGCGTCCGGGCGCGCTGGAGGACGCGGAGGAGCCC T ACA **%** 25 CHG GHG F TTC Y V GTC TACT နှ Tcc ပည္သိုင္သ R AGG ၁ ဥ F Y O CAG GAG I ATA V GTG O CAG 9 9 9 0 Y TAT F Y TAT I ATA N AAT ပိုင္ပင္ပ L TTG ၁၁၅ A GCT æ F TTC Y TAT နှင့် မြင့်င GG GGA GGA P CCG A GCA LCIG ACC A GCC CHC ဂ ညီ ი ი ტ ტ PCCA နှင့် င TGC P CCA P M W CHC န TCC Y TAT ۳ ۲ F EGAG CAG r CTG GIC V GTG LCTT E GAG CAG N AAT ი მემ r CHG GGA V GTG I ATC C TGT A GCC CIG A GCG E GAA ၁ ၁၅ ဗ ဗ္ဗ G GGA M ATG R AGG P CCG PCCA နှင့် TCC ATG Y TAT GG GG ပ္ပ္ဆိုင္သ Σ ი მგი CTG က္ခ ¥ TGG မှ င်င်င

Fig. 4A

166 632 1048 1206 890 696 1285 1364 1443 1522 1680 1127 1601 CTCTCCAGGGCATTCTCAGGCCCCGGGGGTCTCCTTCCCTCAGGCAGCTCCAGTGGTGGGTTCTGAAGGGTGCTTTCAAA TGCACTGACCATGTTGTCATAATTAGAATAAAGAAGTGGTCGGAAATGCACATTCCTGGATAGGAATCACAGCTCA TGCTGGAGGTGCAGGTGGCATGTAGAGGGGCCAGGCCGAGCATCCCAGGCAAGCATCCTTCTGCCCGGGTATTAATAGG CTTCCTGCCCCAAACTGAGACATTGCATTTTGTGAGCTCTTGGTCTGATTTGGAGAAAGGACTGTTACCCATTTTTTG TGTGTGAACGCTGACCTGTCCTGTGTGCTAAGAGCTATGCAGCTTAGCTGAGGCGCCTAGATTACTAGATGTGCTGTAT CACGGGGAATGAGGTGGGGGGTGCTTATTTTTAATGAACTAATCAGAGCCTCTTGAGAAATTGTTACTCATTGAACTGG TGGGGTGCCCACGTGCAAGAGGAGAGAGAGAGGGCCTTTCCCTGGCCTTTCTGTCTTCGTTGATGTTCACTTCCAG CTGTTCATATCCTAAAGATAGACTTCTCCTGCACCGCCAGGGAAGGGTAGCACGTGCAGCTCTCACCGCAGGATGGGGC AGCATCAAGACATCTCATGGAAGTGGATACGGAGTGATTTGGTGTCCATGCTTTTCACTCTGAGGACATTTAATCGGAG GAACGGTCTCGTGGGCTGCTAAGGGCAGTTCCTCTGATATCCTCACAĠCAAGCACAGCTCTCTTTCAGGCTTTCCATGG AAGCCCCATGCCGGCCGCTCAGCCGATGAAGCAGCAGCCGACTGAGCTGAGCCCAGCAGGTCATCTGCTCCAGCTGT AGTACAATATATGAACTCACACTTTGTCTCCTCTGTTGCTTCTGTTTCTGACGCAGTCTGTGCTCTCACATGGTAGTGT V A C P P P A Y C N T P P P Y G GTG GCG CCG CCG CCG TAC TGC AAC ACG CCT CCG CCC CCG TAC V K A K * GTG AAG TAG Q V CAG GTA

Fig. 4B

151 GTCGACCCACGCGTCCGGCGCGCGTCCTTCTGCCGGCTTCAGCTCGTATCCCCGGAGTCCACCCGGCCCGTCCGGGGGT G R V A A L L L G L L V E C T E A K K GGC AGG GTG CTG CTG CTG CTA GTG GAG TGC ACT GAG GCC AAA AAA GCGGACTGGCCCTGAGCTGGCCCGGCTTCGGACGGTCCTCGCTGGAGCC ATG GGC CGC

| 45 271 | 65 331 | 85 391 | 105 451 | 125 511 | 145 571 | 165 631 | 173 655 |
|-------------|-------------|------------------|------------|--------------------|------------|------------------|------------|
| C FGC | FTTC | M ATG | P CCA | CCC | PCCT | Y TAT | ٠ |
| D GAC | A G G | 8 0 0 0 | CCA | ი გ ეე | S TCA | CCC | |
| E GAA | F TTT | ი გ | CAG | ი წეგ | N AAT | CCC | |
| Y TAT | Y TAT | န ၁၅၁ | R AGG | CCT | CCC | CCA | |
| si DD | ¥ G G | I ATT | T ACC | GAC | CAG | PCCT | |
| R G H | CIG | F | Y TAT | ACC | S C | ACG | |
| ပ္ပင္မ | r Agg | F | s TCC | Y TAC | CAG | N AAC | |
| I ATA | CAG | G GGT | V GTG | Y TAT | F HHC | ည သို့ | |
| Y TAT | I ATA | ₽ | N AAT | CCA | A GCT | Y TAC | |
| Y TAC | S TCC | GGT | F TTC | P CCG | M ATG | S | |
| T ACA | CIT | c TGT | T ACA | ත <mark>ද</mark> ව | ₽ GCT | CCT | |
| ССС ССС | ₽ | C FGC | CCC | M ATG | M ATG | CCT | |
| Y TAT | r Agg | F | E GAG | CAA | ACC | CCC | * TAG |
| CIC | V GTG | LCTG | e Gag | CAG | n AAT | PCCA | K AAG |
| . G GGA | C TGT | V GTG | i Att | GCA | ი ე | Y TAC | GAC |
| e gaa | ပ မရင် | G GGT | CIC | 0 66A | V GTT | TACT | k Aag |
| F TTT | r Agg | A A T G | PCCA | PCCA | PCCT | TACA | V GTG |
| Y TAT | s TCC | M ATG | PCCA | AGCT | n AAT | 9 990 | V GTG |
| W TGG | 9 9 9 | CIG | P | CCT | M ATG | G GG A | O CAG |
| ၁ ညီ | C TGT | CHG | Y TAT | N AAT | ტ ტ | H | e gaa |
| | | | | | | | |

Fig. 4H

CAAGATGCTACATCAAAGGCAAAGAGGATGGACAGGCCCTTTTGTTTACCTTCCCATCCTCACGATACTTGCTGAJAG 734

2235 1840 1919 2077 2314 1998 1682 1761 892 TTATGGGCACTATAGAGCTGAGGGCACATTAGGCCGGGTAGTTACATTGACCCTTGGAGAGGAAGAGGACAGCCAAAG AAACTCAGCAAAGCAAGACCAGCATTGCTGAGTTAGAGCTAGGGTTGTATGTGTGATCCCAACAGAGATGTGCTGGCCTCA GAAGAGGGGACGTTTGTGGATAGAGCCGTGAAAACCTACTTAGTTGCACAGATGACATAATCAAAAGTAGAAAAAAGAAA TGTAGTTAGAGATGCCATTTCCCAGGTGAGAATCAGAGCTCATCCATAGATTTACAAGTAGTGGCTGGAGTTAACAGTA TGGAGTTCTTTTCCCTTGCGTAGTTAGTCACGTTGATGTGTATTTAAACCCCAGGTTGAGACCTTGTGTACTAAGAGCAA TAAATGAGAAAATCAGAGCCATTTGATAAACTGTTACTTGTTGGATCAGGCATCCAAAAGTGTCTCTTGAGTGGACATT GAGTATTCTTTACCACCTACAAGACCAGGAGGCATGGTGTCATTCTCCATTGGGGTATTTATATGAGGTAGAGGTTCAG GAATCGACAGTAGCTGTGTGGGCTTAGTTTAAGGACTGAAAGCATAGGGACTGGTAGACAGTTTCATAGGAAACTGCGG GGAAGGAATGGATACCTTTAAAGACAGTTTGTGGATGCAGATGCTGCCACCCATCATTGAGCACCCTTGTGTCTCTGGC TTCCTGTCACTGGATCCAGTACCCCTCCATGCTTGGGTCCTTGTTTTACATAAGACAACAAGACAATGTCTGCTGTT CCTTTTTAAAATCAAGAAGCACAGTCAGAGCTGCCCCTGGGATTGCATCAGGGAACGGCTGATCAAGGCATTCAGTGTC CATGACTAAATCTTATCTTTTGATAGCAAATCCTTTTAAGAAACTGAACAATTGCTAAGGCTCAGCAATTTTATACTC CAGATAAAGTGGTCAGGCTGAGATAAGACTCACATGATGCAGTAGTTGGCAGTGAACTTCGAAGAGACACTATCCACCA TCCCAGCCCCATTCTCCTAATAGAAGCTGTGGGGCTGTGTTGTTGATGCTCTTTGGTCTCCACTCACATTTTGAAAATAG GCTTTCCTCTGCAGGAATAGGAAAGACCCAAGTACATATTTGCTTCCACTTAAAAATGAGGGTCAGAACCAGGCCTCAG GCGTTTTATGAATCATCGTCTGGCTTTTTTTTAGTGCATGTATTGAAGTGAGGGTGTCCTTTGAGATCAGATGGGGAG AGTGAACTCTGCGGGGGGTGGGGTGTCTCTACTCAGAGGGCTCCAACACCCCTTTTCTTAGGTAGTTCTGGTGATGGGTT TACAATCAAGACGACTACATGGTCCAAACATTTCTTCTTCTTCTATCACTTGTGGCTTTAACTTCCATTTCCTCCGTT CAATGTCTGTGTAAGGTAAATTTTTGTTTGCCATTGAGCCCAATTGGAATTCCTTCTGACGTCAACACTGACAATGCCT GGTGGTCCAAGGGAAAACTTGGATATTCTCAAAGCAAGCCCAGCTCTCTTTCAAGTCTTTTGTGGAGGACATTTGAATC TTGGACATCTATAGTTAAATAAGGCCATTAGAGAGGGGAAATCTTTAAGTTAGGGGAAATTCTCTAAATGGAGAGACATT <u> ATGGAAATTGCACTTCTGGGTATATGTCCCAGCATCCTTGTTTTCTTATGTTTGGTGAGTAAGGCTCACCCCTTCCAGC</u> CACACTGTCTCCTCTGTTGCTTCTGTTTCTGATGTAGTCTGTGCTCTCTGAGAGAGTGTGGCAACAGTCCCTGAGGGGTT

Fig. 41

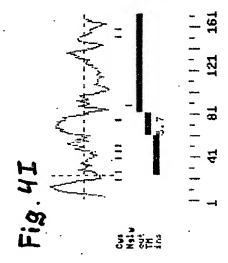
71/96

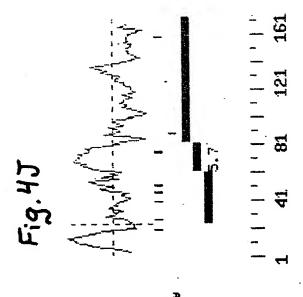
2788 2867 2915 AGCTCTACTTCTGTGTGTGTGTCCTGTAGAGCCGGGGCTTGGGCACAGACATGAGGCAGACTTGTGCATGCTCTTTC TTGGCAACACTTTGGCTCATATTTCTTGTTCTCTTTTGATAGAGTCCTGTTTCCTATGTATTTAAAAAATAAAAGTG

Fig. 40

| | 10 | 20 | 30 | 40 | . 20 | 9 | 70 |
|--------|--|------------|-------------|-------------------|---------------------|-------------|--------|
| Hum. | MRRQPAKVAALLLGLLLECTEAKKHCWYFEGLYPTYYICRSYEDCCGSRCCVRALSIQRLWYFWFLLMMG | LLLECTEAKK | HCWYFEGLYP | LYYICRSYEI | OCCGSRCCVR1 | ALSIQRLWYFW | FLLMMG |
| | ••••••••••••••••••••••••••••••••••••••• | | | | | | •• |
| Mur. | MGRRLGRVAALL | LLVECTEAKK | HCWYFEGLYP | PYYICRSYEI | OCCGSRCCVR | ALSIQRLWYFW | FLLMMG |
| | 10 | 20 | 30 | 40 | 20 | 09 | 70 |
| | 80 | 06 | 100 | 110 | 120 | 130 | 140 |
| Hum. | VLFCCGAGFFIR | MYPPPLIEEP | AFNVSYTROPI | PNPGPGAQQI | PGPPYYTDPG(| 3PGMNPVGNSM | AMAFQV |
| | | | •• | •• | •• | | •• |
| Mur. | VLFCCGAGFFIRRRMYPPPLIEEPTFNVSYTRQPPNPAPGAQQMGPPYYTDPGGPGMNPVGNTMAMAFQV | MYPPPLIEEP | TENVSYTROPI | PNPAPGAQON | IGPPYYTDPG (| SPGMNPVGNTM | AMAFOV |
| - - | 80 | 06 | 100 | 110 | 120 | 130 | 140 |
| | 150 | 160 | 170 | | | | |
| Hum. | PPNSPQGSVACPPPPAYCNTPPPPYEQVVKAK | PAYCNTPPPP | YEQVVKAK | | | | |
| | | | ••• | | | | |
| Mur. | Mur. QPNSPHGGITYPPPPSYCNTPPPPYEQVVKDK | PSYCNTPPPP | YEQVVKDK | | | | |
| | 150 | 160 | 170 | | | | |

Fig. 4H





GTCGACCCACGCGTCCGCAGCTTTGGACACTTCCTCTGAGGACACCTTGACTAACCTCCAAGGGCAACTAAAGGA

| TCAAGAAAGCCCCCACCACCACCACCACACACATCACTCCTCCTCCAGGAGA TTG TGT ACA AAG ACA ATC TCAACAAAACCCCCCACCCACCACCACACACATCACTCTCTCTCCTC | 6 150 | 26 210 | 46 | 930 330 | 86 390 | 106 450 | 126 510 | 146 570 |
|---|----------|-----------|------------------|------------|-----------|------------------|------------|------------|
| N | υ | I ATT | G GGA | EGAG | S AGT | LCTA | V GTT | ATG |
| N | | | A GCT | S TCT | I ATC | ₽ GCG | F TTT | E GAA |
| N | | CAG | CAA | G GGT | K AAA | K AAA | CII | N AAT |
| N | | S TCT | V GTT | S AGC | I ATA | I ATC | PCCA | L TTA |
| N | | s TCC | a G T | L TTA | N AAT | G GGA | S TCT | N AAC |
| N | | S TCA | Y TAT | D GAT | s TCA | V GTG | e Gag | K AAG |
| AGAAAGGCCCAGCACAGAAA V L W G C GTC CTC TGG GGA TGT T CCT GGA ATC AAG GCA A AAG ATG ATT GAG CAA A CTT GAA TTT CTA AAA G TTT TCA TTT CCA AAT A AAC CAT GGC ACT GCC A TAT AAC TCC TTT GCT G | | QTC | D GAC | CCA | FTT | G GGA | FITC | L TTA |
| AGAAAGGCCCAGCACAGAAA V L W G C GTC CTC TGG GGA TGT T CCT GGA ATC AAG GCA A AAG ATG ATT GAG CAA A CTT GAA TTT CTA AAA G TTT TCA TTT CCA AAT A AAC CAT GGC ACT GCC A TAT AAC TCC TTT GCT G | :AGG? | Y TAT | CHT | CIC | N AAT | CCT | ტ მტტ | I ATT |
| AGAAAGGCCCAGCACAGAAA V L W G C GTC CTC TGG GGA TGT T CCT GGA ATC AAG GCA A AAG ATG ATT GAG CAA A CTT GAA TTT CTA AAA G TTT TCA TTT CCA AAT A AAC CAT GGC ACT GCC A TAT AAC TCC TTT GCT G | CTGC | CIC | A GCA | K AAA | Y TAC | V GTG | ¥ TGG | P CCC |
| AGAAAGGCCCAGCACAGAAA V L W G C GTC CTC TGG GGA TGT T CCT GGA ATC AAG GCA A AAG ATG ATT GAG CAA A CTT GAA TTT CTA AAA G TTT TCA TTT CCA AAT A AAC CAT GGC ACT GCC A TAT AAC TCC TTT GCT G | GCTC | n AAT | r Agg | K AAG | NAAAC | 7 7 7 7 | D GAC | k Aaa |
| AGAAAGGCCCAGCACAGAAA V L W G C GTC CTC TGG GGA TGT T CCT GGA ATC AAG GCA A AAG ATG ATT GAG CAA A CTT GAA TTT CTA AAA G TTT TCA TTT CCA AAT A AAC CAT GGC ACT GCC A TAT AAC TCC TTT GCT G | TCTA | T GG | Q CAG | E GAA | V GTA | A GCT | TACA | e GAG |
| AGAAAGGCCCAGCACAGAAA V L W G C GTC CTC TGG GGA TGT T CCT GGA ATC AAG GCA A AAG ATG ATT GAG CAA A CTT GAA TTT CTA AAA G TTT TCA TTT CCA AAT A AAC CAT GGC ACT GCC A TAT AAC TCC TTT GCT G | TGGA | L CTG | ACT | K AAA | Y TAT | 1 11 9 | SAGC | M ATG |
| AGAAAGGCCCAGCACAGAAA V L W G C GTC CTC TGG GGA TGT T CCT GGA ATC AAG GCA A AAG ATG ATT GAG CAA A CTT GAA TTT CTA AAA G TTT TCA TTT CCA AAT A AAC CAT GGC ACT GCC A TAT AAC TCC TTT GCT G | cago | CHC | I ATT | L | D GAT | s TCA | ATC | G P CCC |
| AGAAAGGCCCAGCACAGCAGG V L W G C GTC CTC TGG GGA TGT CTT GGA ATC AAG GCA AAG ATG ATT GAG CAA L E F L K CTT GAA TTT CTA AAA TTT TCA TTT CCA AAT N H G T A AAC CAT GGC ACT GCC TAT AAC TCC TTT GCT | AGAT | TIC | r Agg | M ATG | V GTT | T ACC | N AAC | E GAG |
| TCAAGAAAGGCCCAGCACAG P V L W G CCA GTC CTC TGG GGA Y P G I K TAC CCT GGA ATC AAG ATG AAG ATG ATT GAG S L E F L TCT CTT GAA TTT CTA A F S F GCC TTT TCA TTT CCA T N H G TT ACC AAC CAT GGC ACT L Y N S F CTG TAT AAC TCC TTT | CAGE | | | CAA | K AAA | N AAT | ₽ | A GCT |
| TCAAGAAAGGCCCAGC P V L W CCA GTC CTC TGG Y P G I TAC CCT GGA ATC M K M I ATG AAG ATG ATT S L E F TCT CTT GAA TTT A F S F GCC TTT TCA TTT T N H G ACC AAC CAT GGC L Y N S CTG TAT AAC TCC | ACAG | G GGA | K AAG | e Gag | L | CCA | TACT | FTT |
| TCAAGAAAGGCC P V L CCA GTC CTC Y P G TAC CCT GGA A K M ATG AAG ATG S L E TCT CTT GAA T N H ACC AAC CAT L Y N CTG TAT AAC CTG TAT AAC | CAGC | 1 G € | I ATC | I ATT | F | F TTT | ဗ ဗ္ဗ | s TCC |
| TCAAGAAA P V CCA GTC Y P TAC CCT A K ATG AAG S L TCT CTT A F GCC TTT T N A CT T N A CT T N A CT T N A CT T CT | ၁၁၅၅ | CHC | 6 66 A | M Atg | e gaa | S TCA | H CAT | N AAC |
| TCAA TAC TAC TAC TAC TAC TAC TAC TAC TAC | GAAA | | CCT | K AAG | CTT | F | N AAC | Y |
| | TCAA | | | M ATG | | ₹ | ACC | CHG |

Fig. 5A

75 / 96

186 690 206 750 226 810 246 870 266 930 L CHG E r GFC L ი მ CAA M ATG A GCT ACC. N AAC M ATG S FTTC P CCA E GAA s TCC T ACA NAAC F E 3AA SICT G P ACA N AAT F CAG s AGC N AAC CAA L CTG S AGT r. CFC CCA s AGC H V GTT s TCC ဗ ဗ္ဗဗ T ACA IATC R GGC A GCT r Trg PCCA နှ TCC N AAC LCTA Y TAC F TTT **A** GCC GAA H IATC CAA S TCC F F TTT Y TAC K AAG P CCA S N AAC L Y V GTA I ATC PCCC LCTA DGAT G GGT LCTC A GCG S TCC N AAT I ATC E GAG CAA **A** GCG K AAG or c STCT I ATT L e Gag A GCA I ATA AAA AAA L L TTG F TTT K AAA CIC V GTC N AAC FTT E GAA I ATT မှ ပိုင် PCCT TACT L ACC R CGG V GTT FITC PCCT Y TAC N AAC GAC S TCC S TCC EGAG CCA Y TAT S AGT CHC CHC A GCA DGAC s ICA E GEC CHT T ACA I ATT F TACT V GTG န TCC I ATT Y A A GCC L AAC **A** 300 I ATT K AAG N AAC CCC CCC I ATC M ATG ACC. r CCC CC ရ ၁၅၅ I ATC EGAG GGA N AAT CCT L ITA TACT D GAC I ATT F CHI R AGG ATC V GTT ລ ດີດີດ v GTG CHC TACC Y V GTT I ATT

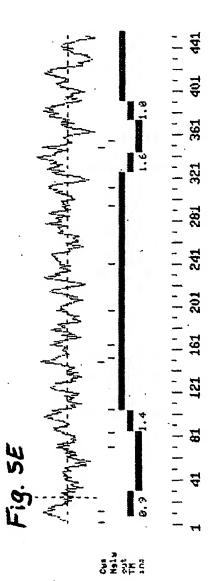
Fig. 51

76/96 426 446 456 1470 1500 Y TAT F S AGC N AAC PCCA I ATA K AAG s AGC L L CTC CIG CA S L ස ප V GTC N AAC 0 9 8 G GGA D GAC N AAT K AAA CTG $_{
m TTG}$ CAC T ACC S F TTT S TCC GGT GGT H г ССА e Gag I ATT E GAA N AAT CCA L က ည L TTG 75 ¥ € CTG r Tr I ATT V GTA S TCC L LCTT A GCT ရ ၁၅ ၁၅ H V GTT CIT s TCG CCT FITC FTC L CTA S AGT SGC CGC F TTT G GGT TGA CCT S AGT FTTC I ATT G GGA E A GCC ССА В N AAT CIT R AGA CAA E GAA STCA N AAC CAG CAG V GTT K AAG V GTT L. CTG L TTG CAG GAA FITI ი მი FITC s TCT R AGG X AAA I ATT K AAG R AGG s TCA GAC L ITG A GCA DGAT L TTG W TGG န TCC A I G V GTC N AAT S S TCA R Q AGA CAG N AAT ACA, E GAG ၁ ၁ ၁

1579 1658 TTTGTTTGGTTTGGGGCAAGAATTCTAGGACAAGAGCTAGGCATGTACTTCTGACCAGGTGGGTAAGCAACTCTAAG TTGCCGGTTTGCAATTCACCCCAGGAAGTAAATGGTCCTTAATCCTACAACTACTGTAAACCCAGAAGGGGAAAGACAGT

Fig. 5(

Fig. 5D



| 10 20 30 40 50 60 286 MCTKT-IPVLWGCFL-LWNLYVSSSQTIYPGIKARITQRALDYGVQAGMKMIEQMLKEKKLPDLSGSESL : : : : : : : : : : : : : : : : : : : | 70 80 100 110 120 130 286 EFLKVDYVNYNFSNIKISAFSFPNTSLAFVPGVGIKALTNHGTANISTDWGFESPLFVLYNSFAEPME | 140 150 286LKN-LNEMLCPIIASE ::: :: :: :: :: :: :: :: :: :: :: :: :: | 160 170 180 190 200 210 220 286 VKA-LNANLSTLEVLTKIDNYTLLDYSLISSPEITENYLDLNLKGVFYPLENLTDPPFSPVPFVLPERSN : .: .: .: :: :: :: :: .: .: :: :: .: .: |
|--|--|---|---|
| 2 B | 2 B | 28 | 28 |
| B B | B B | BE | BB |

| | 230 | 240 | 250 | 260 | 270 | 280 | 290 |
|-----|--|--------------------------------------|--|--------------------|--------------------|------------------------------|---------------|
| 286 | SMLY | SASFAHFTAGV | FNLTLSTEE | ISNHFVQN | SQGLGNVLSRI | AEIYILSQPFN | IVRIMA |
| 0 | : | | STATEMENT STATEMENT TO STATE STATEMENT STATEME | | THKFFGTFT, PRVAK | * AKKFP-NMKIC | : |
| 4 | 280 290 | 300 | 310 | .320 | 330 | 340 | |
| | 300 | 310 | 320 | 330 | 340 | 350 | 360 |
| 286 | TEPPII | TLDIPASIMME | TQPKNSTVE | TIVSMDFVAS | TSVGLVILGQR | LVCSLSLNRF | KLALPE |
| BPI | . :::: :: :: :: | : TEYPAVDVQAFAV | AVLPNSSLASLI 380 | SLFLIGMHTTO 390 | : | :: ::: LVGELKLDRL1 410 | : : :LELKH |
| 286 | 370 380 390 400 410 420 430 430 390 SNRSNIEVLEGFLLISTDLKYETSSK | 380 IILSSILHEGVL | 390 PLANAKLOOC | 400 SFPLPNPHKE | 410 LFVNSDIEVLE | 7 420 GFLLISTDLK | 430 KETSSK |
| BPI | | ::.::::::::::::::::::::::::::::::::: | PRVNEKLOK | SFPLPTPARV | OLYNVVLOPHC | NFLLEGADVV | : YK |
| | 420 430 | 440 | 450 | 460 | 470 | 480 | |
| 286 | 440 450 OOPSFHVWEGINLISROWRGKSAP | 450 IISRQWRGKSAP | | | | | |
| | | l | | | | | |
| BPI | | 11111111111 | | | | | |

| 10 20 60 60 286 MCTKTIPVLWGCFLLWNLYVSSSQTIYPGIKARITQRALDYGVQAGMKMIEQMLKEKKLPDLSGSESL. : | : . : : : : : : : : : : : : : : : : |
|--|-------------------------------------|
| 286 MCTKTIPV : | : .:. RENP VSSKLQPY 210 |

Fig. 51

RMVYLGLSDYFFNTAGLVYQEAGVLKMTLRDDMIPKESKFRLTTKFFGTFLPEVAKKFP-NMKIQIHVSA

TEPPIINLQPGNFTLDIPASIMMLTQPKNSTVETIVSMDFVASTSVGLVILGQRLVCSLSLNRFRLALPE

340

330

320

310

300

286

RENP

RENP

286

STPPHLSVQPTGLTFYPAVDVQALAVLPNSSLASLFLIGMHTTGSMEVSAESNRLVGELKLDRLLLELKH

SMLYIGIAEYFFKSASFAHFTAGVFNLTLSTEEISNH--FVQNSQGLGNVLSRIAEIYILSQPFMVRIMA

260

240

270

SNRSNIEVLRFENILSSILHFGVLPLANAKLQQGFPLPNPHKFLFVNSDIEVLEGFLLISTDLKYETSSK RENP SNIGPFPVELLQDIMNYIVPILVLPRVNEKLQKGFPLPTPARVQLYNVVLQPHQNFLLFGADVVYK-400 410 400 390 286 QQPSFHVWEGLNLISROWRGKSAP 286 RENP

GTCGACCCACGCGTCCGGGGAATTGCAGCAGGAAATATGTGAAGAGTTTTTAAACCCACAAATTCTTTACTTTAGA

83 / 96

389 149 68 329 108 449 28 209 269 128 509 148 Q CAG I ATT ¥ TGG CAA AAT E GAA T ACA z CAG $_{
m F}$ TACT ATG K AAG ₩ TGG V GTG M ATG A GCA k AAG GAC FIC AAC I ATA Y TAT V GTC S AGC GCA CCT F TTT **&** 606 E GAA CAA GAA A GCT G GGT Gid E T GAA ACC 闰 Y V GTG G GGT ₽ GCT CCA GAA K AAG GAC G GGA GAT L CŢĀ ATTAGTTGTTACATTGGCAGGAAAAATAAATGCAGATGTTGGACC ATG TTG Н I ATT E GAG ဗ ဗ္ဗ A GCA V GTT CAC 耳 Σ LCTG ng TgT r CIG AAA A GCT R CGA LCTA CIT က္ရ ည ဗ ဗ္ဗ CGA CCT IATT K AAA TACT Y TAT H TCI I ATT FITC E . M GAA ATG ဗ္ဗဗ္ဗ CCA R AGG CAG ဗဗ္ဗ CAA N AAC CHG CHG M ATG CAT L TTA H S AGC AAC V GTT 工 z R AGA CAA ∨ GTG GTA STCT N AAT U H I ATC GTG AGG s TCA N AAC CTT > α I ATC s TCA N AAT I ATC CCT ပ္သင္ပ s AGC V GTC V GTG GAA GAA Y TAT R AGG CIG AAC Z N AAT S AGT ი მმმ TCC GAT G ĞĞT

Fig. 6A

N AAT

X AAA

ACC.

E GAG

S AGT

ი მმმ

₩ TGG

DGAC

FTT

A GCA

CHC

E GAA

a G

s ICT

N AAT

V GTG

84 / 96

248 869 308 1049 188 689 208 749 268 929 288 989 228 809 CAG I ATT ¥ TGG I ATA უ ცე M ATG E TTT R CGA Y TAT CAG CAC s AGC AAA AAA Q CAG CIG M ATG T ACC D GAT A SCA STCA IATC LCIA N AAC Y TAT K AAA G G G LCTT I ATT F TTT M ATG N AAT ·L CAG PCCC I ATT ၁၉ E V GTG V GTC AGCT s AGC K AAA N AAT CAA CAG LCTG K AAA K AAA N AAC V GTG F TTT Y TAT A GCA T ACC S R AGG Y FAT E ရ ရ 9 9 9 9 N AAC A GCT I ATC TACA CCA H E TTT C HGH M ATG r Trg FITC GGA K AAG M ATG K AAG LCTT T ACC Y A GCT E E V GTT G GGA G GGA GAT s TCC K AAG I ATT G GGT CTT CAG TACT I FITI A GCC V GTT L Y IAT ი ემ ACT S AGT ACG I ATA M ATG A GCA LCTT L H F TTC M ATG I ATT F CCC 3 3 3 C C K AAA CAA L TTA A GCT Q CAG AGCA M ATG F DGAT R AGA AGCT r GG IATC L lTG ဗ္ဗဋ္ဌ CIC Y TAT L TTA PCCA N AAT M ATG AGCT L FTT V GTA I ATT ACC L TTG SAGT F LTT R AGA s AGT 耳耳 ACC L TTG T ACC CTGT Y

Fig. 6F

| 348 1169 | 368 1229 | 388 1289 | 408 | 424 1397 |
|-------------|-------------|-------------|------------------|-------------|
| PCCT | CIG | D GAT | Q CAG | |
| GHC < | M ATG | V GTG | A A T G | |
| ACG | A A A A | H | J D P D | |
| MATG | V GTG | A GCT | H CAT | |
| D GAT | GAC | ¥ TGG | I ATC | * TGA |
| r Aga | E | E GAA | I ATC | L |
| V GTC | PCCA | PCCT | e gaa | V GTA |
| R AGA | N AAT | I ATT | N AAT | A GCC |
| Y TAC | S TCA | N AAT | Y TAC | GAG |
| R AGG | CIT | K AAG | M ATG | C TGT |
| V GTA | ¥ TGG | H CAT | R CGT | ი გე |
| PCCT | D GAC | Y TAC | HCAC | G GGA |
| TACT | CAG | IATC | CCT | CAG |
| 면 (A). | G GGT | CIC | A GCT | S TCC |
| Q | ი გეგ | N AAC | D GAT | L |
| N AAT | TACA | ACC | TIG | N AAC |
| ပ မှု | ¥ TGG | ∨ GTG | ი იი 1 | H ACC |
| K AAA | M ATG | e GAG | W TGG | E GAG |
| e gaa | A GCA | S TCT | I ATC | e Gag |
| L CTG | ACA | CHC | H H H C | CAG |

1713 1476 1555 1871 2029 1634 1792 1950 **TTAAAGTACTTATTAGGTAAATAGAGGTTTTTGTATGCTATTATATATTCTACCATCTTGAAGGGTAGGTTTTACCTGAT** CAACAACAAAATCAGTGTTACAGTATGGATGAAATCTATGTTAAGCATTCTCAGAATAAGGCCAAGTTTTATAGTTGCA **AGCATCTGACACTGACGATCTTAGGACAACCTCCTGAGGGATGGGGCTAGGACCCATGAAGGCAGAATTACGGAGGCA** ATTTTGGAGCACTAAAGTAAAATGGCAAATTGGGACAGATATTGAGGTCTGGAGTCTGTGGATTATTGTTGACTTTGA CAAAATAAGCTAGACATTTTCACCTTGTTGCCACAGAGACATAACACTACCTCAGGAAGCTGAGCTGCTTTAAGGACAA AAAAAGGGGGGCGC

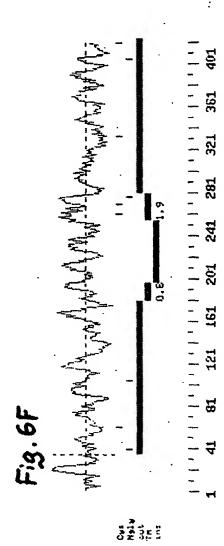
Fig. 60

| 20 30 40 50 60 70 MEMWLLILVAYMFQRNVNSVHMPTKAVDPEAFMNISEIIQHQGYPCEEYEVATEDG . : :: :: ::::::::::::::::::::::::::: | 80 100 110 120 140 YILSVNRIPRGLVQPKKTGSRPVVLLQHGLVGGASNWISNLPNNSLGFILADAGFDVWMGNSRGNAWSRK :::::::::::::::::::::::::::::::::::: | 150 160 200 210 HKTLSIDQDEFWAFSYDEMARFDLPAVINFILQKTGQEKIYYVGYSQGTTMGFIAFSTMPELAQKIKMYF : :::::::::::::::::::::::::::::::::: | ALAPIATVKHAKSPGTKFLLLPDMMIKGLFGKKEFLYQTRFLRQ-LVIYLCGQVILDQICSNIMLLLGGF::::::::::::::::::::::::::::::::: |
|--|---|--|--|
| 60 GYPCEEYEVATED ::: ::::::: GYPNEEYEVVTED 50 | NSRGN ::::: NSRGN 120 | ELAQK :: SLAKR 190 | O SNIMLI SNALF |
| 60 QGYPC ::: WGYPN | 130 DVWMG::::: DVWLG | 200 FSTMPI | 270 LDQICSI ::: LNLLCSI |
| EIIQH : 2MITY 40 | ADAGE:::. ADAGY 110 | MGFIA .::: IGFIA 180 | OCGQVI |
| SO AVDPEAFMNISEIIQHQ : :::.: PGSPEVTMNISQMITYW 30 | 120 LGFILL | 190 SQGTTI | 260 LVIYLCC :: LATEVC |
| VDPEA. | LPNNS: | YYVGY ::: HYVGH 170 | FLRO-J FFDQF 240 |
| 40 MPTKA | 110 NWISN: | 180 GQEKI :: GQKQLI | 250 LYQTR] :- |
| VNSVH : FGKLH | VGGAS :. LASAT 90 | ILOKT :: IVKKT 160 | GKKEF: GDKIF |
| MFQRN TIHGL | 100 LOHGL ::::: LOHGL | 170 AVINE :.:: ATIDE | 230 240 250 260 270 TKFLLLPDMMIKGLFGKKEFLYQTRFLRQ-LVIYLCGQVILDQICSNIMLLL(.::.::::::::::::::::::::::::::::: |
| ILVAY : : ISVLG | RPVVL::::: | RFDLP ::: KYDLP 150 | LPDMM .: VPQSL 220 |
| 20 MEMWLLII . : :: LTMASLIS | 90 120 140 2PKKTGSRPVVLLQHGLVGGASNWISNLPNNSLGFILADAGFDVWMGNSRGNAWSRK: ::::::::::::::::::::::::::::::::: | 160 170 180 190 200 2 SYDEMARFDLPAVINFILQKTGQEKIYYVGYSQGTTMGFIAFSTMPELAQKIKM ::::::::::::::::::::::::::::::::::: | 230 TKFLL .:. |
| VSHRM .LL | | 0 EFWAFS :::::: EFWAFS 140 | AKSPG .:: TKSLI 210 |
| 10 SRQWIVS : | 80 YILSVNRIPRGLV(::: :::: : YILEVNRIPYGKKI | 150 (IDQDE) : : | 220 ALAPIATVKHAKS :::::::::::: ALAPVATVKYTKS 200 |
| 10 MLETLSRQWIVSHR : MWLL | | 150 160 170 180 190 200 210 294 HKTLSIDQDEFWAFSYDEMARFDLPAVINFILQKTGQEKIYYVGYSQGTTMGFIAFSTMPELAQKIKMYF . :::::::::::::::::::::::::::::::::::: | |
| 294 HLP | 294 HLP | 294 HLP | 294 HLP |

Fig. 6E

294 EAVL

HLP ----



| 09 | AMEMWLLILVAYMFORNVNSVHMFTKAVDFEAFMNISEIIQHQGYFCEEYEVATE :: ::. : :: :: :: ::: ::: ::: ::: ::: | 50 | DGYILSVNRIPRGLVQPKKTGSRPVVLLQHGLVGGASNWISNLPNNSLGFILADAGFDVWMGNSRGNAWS | DGYILCLNRIPHGRKNHSDKGPKPVVFLQHGLLADSSNWVTNLANSSLGFILADAGFDVWMGNSRGNTWS 60 110 120 | 200 | 294 RKHKTLSIDQDEFWAFSYDEMARFDLPAVINFILQKTGQEKIYYVGYSQGTTMGFIAFSTMPELAQKIKM | LAL RKHKTLSVSODEFWAFSYDEMAKYDLPASINFILNKTGOEOVYYVGHSOGTTIGFIAFSOIPELAKRIKM | 190 | 270 | KSPGTKFLLLPDMMIKGLFGKKEFLYQTRFLRQLVIYLCGQVILDQICSNIMLLLGG | FFALGPVÁSVAFCTSPMAKLGRIPDHLIKDLFGDKEFLPQSAFLKWLGTHVCTHVILKELCGNLCFLLCG 200 210 220 230 230 240 250 260 |
|----|--|-----|--|--|-----|--|--|-----|-----|---|---|
| 50 | SAFMNISELLO :- ::::: ETNMNVSELLS | 120 | SLGFILADAG | SSLGFILADAG | 190 | SYSQGTTMGFI | SHSOGITIGEL | 180 | 260 | NEVIYEGGVI: | ILGTHVCTHVI 250 |
| 40 | AFTKAVUPE SCKLTAVDPE | 30 | SNWISNLPNN | SSNWVTNLANS | 180 | KTGQEKIYYVG | THEOEOVYYVC | 170 | 250 | SFLYQTRFLRÇ | FLPOSAFLKW 240 |
| 30 | AFORNVNSVH .VIWPLHSEGS | 20 | /LLQHGLVGG | /FLQHGLLADS | 170 | PAVINFILOE | PASINFILME | 160 | 240 | MIKGLFGKKE | ILIKDLFGDKE 230 |
| 20 | EMWLLILVAYMFORNVNS :: : | 10 | DPKKTGSRPV | NHSDKGPKPVV 80 | 160 | FSYDEMARFDI | SYDEMAKYDI | 150 | 230 | GTKFLLLPDN | PMAKLGRLPDF 220 |
| 10 | SKCW LVSHKM | 08 | SVNRIPRGLV | CLNRIPHGRK | 150 | SIDQDEFWA | RKHKTLSVSODEFWA! | 140 | 220 | KHA | PVÅSVAFCTSI 210 |
| | Z94 MLETLSKÇWIVSH : IAL M | 70 | 294 DGYILS | LAL DGYILC 60 | 140 | 294 RKHKT | LAL RKHKTI | 130 | 210 | 294 YFALAPIATV | LAL FFALGI 200 |
| | | | | | | | | | | | |

Fig. 6G

Fig. 6H

K AAG

E GAG

A GCT

CTA

Y TAC

r TTG

V GTG

CCT

D GAC

CCA

CIG

ი მი

K AAG

E GAG

L CTG

102 375 122 435 142 495 CAG E GAG L F F F F C s AGC GAC IATC L TTA V GTC L. CTG N AAT GAG 田 M ATG s AGC V GTC CAG ი ი L GAG 团 PCCG ACC ATC ဗ္ဗဗ္ဗ I ATT CAG AGCT F V GTG v GTG DGAT F TTC CAG Y TAT T ACC STG TACG s TCT A GCT V GTG AAC Z CCA ი იცი **8**600 T GG s AGC P CCC CCC GAG 臼 K AAG LCTG CHT E GAG I ATC ACC ACC GGT ტ CCC CCC A GCA CIG ₽ GG ი გვვ s TCT CIG ඉදු ACT W TGG s AGT E GAG T ACA CGC ρζ A GCT F CIG F TTC s TCT CIC TGG 3 Y TAT CIG F N AAT s AGT T ACA ACC H FTTC R AGG I ATC IATC V GTG F TTC TIC ပိုင္ပင္ပ M ATG ACG A GCT A GCC N AAC GAG 回 F TTC AAG I ATC K AAG V GTC CIG GAG 臼 T ACA GGA GGA I ATC Y TAC I ATC G GGA N AAC H AGC R CGG GCA s TCA TAC G GGT ď IATT 300 AGCT ACA L CIG TIG GGG AAC ဗ ဗ္ဗ ი მიმ ტ TACT ACC CCT

Fig. 7A

92/96

262 855 282 915 322 1035 302 975 222 735 242 795 V GTG S TCA 3 3 5 5 7 V GTG CIC V GTG FTTC GHC AGCT C TGT A PCCT LCTG S ICT GGA L CHG n Igi SCA K AAG G P D GAT AGCT L CHG GGT ACC. M ATG A L GAA L I ATT Y န ICC L ITG ဗ္ဗဗ္ဗ PCCT L CTĠ R AGG E GAG D GAC CHC L GGA TACA H CAG H PCCT A A A A MATG O CAG T ACC G P CCT နှ Icc V GTG F CIG S AGT GGC CGC L TTG CAG. K AAG N AAT ₩ TGG H L H ATC e Gag д ССС A GCA T ACA M ATG A GCC 9 9 9 9 CTGT T ACG I ATC AGG L CTG S AGT E GAG Y FAC L CCC D GAC o CAG A H M ATG LCTG S . ¥GG T ACC F **A** G P SCT T n TGT ¥ ∏GG L TTG GHC D GAC **A** V GTG A H G Y ပ ဦ L GAA LCTA GEC M ATG S TCA CCT A GCG STCC A GCA DGAT ပ ပို့ F Y T ACA ი მმ A.T.G ж 366 K AAG C IGT G G G G A H A GCT V. GTG Y TAC ACC A GCA H S M ATG CIG S AGT ж С С С PCCA V GTG G GGT s AGC ရ ၁၅ Q CAG P CCC STCC TACT Y TAT ¥ € HCAT L N AAC r AGA V GTA S AGC TCTA FTTC F L L L CFG

Fig. 71

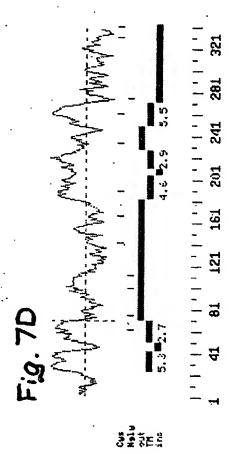
344

1101

93/96

1338 1180 1496 1733 1970 2049 1417 1575 1654 1812 1891 2128 CATTCCTCCCCGTGGAGGCCACCTGGACTTCCAGTCTGGCTCCAAACCTCATTGGCGCCCCCATAAAACCAGGAACTG AAAAAACAAAACAAAAACAAAAAAGCCCTAAGGGACTGAAGAGATGCTGGGCCTGTCCATAAAGCCTGTTGCCATGATAAG GCCAAGCAGGGGCTAGCTTATCTGCACAGCAACCCAGCCTTTCCGTGCTGCCTTGCCTCTTCAAGATGCTATTCACTGA AACCTAACTTCACCCCCATAACACCAGCAGGGTGGGGGTTACATATGATTCTCCTATGGTTTCCTCTCTCATCCCTCGGCA CCCTCAGGGTGGCTGTTACCAGACACCCAGCACCAATCTACAGACGGAGTAGAAAAAAGGAGGCTCTATATACTGATGTT CCICITGITITCCITITICCTGGGITCCTTTTGITCTTCCTTTACTTCTCCAGCTTGTGTGGGCCTTTTGGTACAATGAA **AGACAGCACTGGAAAGGAGGGGAAACCAAACTTCTCATCCTAGGTCTAACATTAACCAACTATGCCACATTCTCTTTGA** GCTTCAGTTCCCAAATTTGCTACATAAGATTGCAAGACTTGCCAAGAATCTTGGGATTTATCTTTCTATGCCTTGCTGA CACCTACCTTGGCCCTCAAACACCACCTCACAAGAAGCCAGGTGGGAAGTTAGGGAATCAACTCCAAAACGCTATTCCT CGTCCCCCTGGGGTGGTTCAGAAGATGCTAGCTCTGGTAGGGTCCCTCCGGCCTCACTAGAGGGCGCCCCTATTACTC

Fig. 7C



| | 10 | 20 | 30 | 40 | 50 | 09 | 70 |
|-----|--|--------------------|---------------------------------------|---------------------------|------------|---------------------|-----------|
| 296 | 296 MATLGHTFPFYAGPKPTFPMDTTLASIIMIFLTALATFIVILPGIRGKTRLFWLLRVVTSLFIGAAILAV | AGPKPTFPMDTTI | ASIIMIFLTAL | ATFIVILPGI | RGKTRLFWLL | RVVTSLFIGA | ILAV |
| CRP | : | NI | ::::.: .:.::.::.:.:.:.:.:: | IAYILILPGV | : .: | : :::::: | LIAS |
| | α | 10 | 50 | 30 | 40 | 50 | ر. د |
| 296 | 296 NFSSEWSVGQVSTNTSYKAFSSEWISADIGLQVGLGGVNITL | STNISYKAFSSEW | ISADIGLQVGL | GGVNITL | TT TETPOO | - XN | EFTW |
| CRP | CRP LIYPCWASGSQMIYTQFRGHSNERILAKIGVEIGLQKVNVTLKFERLLSSNDVLPGSDMTELYYNEGFDI 60 70 80 120 | IYTQFRGHSNER | TQFRGHSNERILAKIGVEIGLQKVNVTL 80 90 | :::: OKVNVTLKFE 100 | RLLSSNDVLP | SSNDVLPGSDMTELYYNEC | : GFDI |
| | 140 | 150 | 160 | 170 | 180 | 190 | 200 |
| 296 | 296 RLGENYAEECAKALEKGLPDPVLYLAEKFT-PRSPCGLYRQYRLAGHYTSAMLWVAFLCWLLANV-MLSM | KALEKGLPDPVLY | LAEKFT-PRSP(| CGLYRQYRLA | GHYTSAMLWV | AFLCWLLANV- | MLSM |
| | • | ••• | | ••• | ••• | •• | :: |
| CRP | SMAEALH | IGLENGLPYPMLS | VLEYFSL | FDWGRHYRVA | GHYTHAAIWE | AFACWCLSVVI | MLFL |
| | | 140 ISO | 09T | 0/.1 | 180 | 0 5 7 | |
| | 210 | 220 | 230 | 240 | 250 | 260 | 0.0 |
| 296 | 296 PVLVYGGYMLLATGIFQLLALLFFSMATSLTSPCPLHLGASVLHTHHGPAF- | TGIFQLLALLF | SMATSLTSPCP | LHLGASV | THTHHGPAF- | WITLTTGLLCVL | LCVL |
| | ••• | • | • | | •• | ••• | ••• |
| CRP | CRP PHNAYKSILATGISCLIACLVYLLLSPCELRIAFTGENFERVDLTATFSFCFYLIFAIGILCVL | MIGISCLIACLVY | LLLSPCE | LRIAFTGENF | ERVDLTATES | FCFYLIFAIGI | LCVL |
| | 200 | 210 | 220 | 230 | 240 | 250 | 260 |

Fig. 71

330 --SPEEGGLLSPRY--RSMADSPKSQDIPLSEAS CRP CGLGLGICEHWRIYTLSTFLDASLDEHVGPKWKKLPTGGPALQGVQIGAYGTNTTNSSRDKNDISSDKTA 310 CRP GSSGFQSRTSTCQSSASSASLRSQSSIETVHDEAELERTHVHFLQEPCSSSST 300 300 ---EAHPKDPD-340 296 LGLAMAVAHRMQPHRLKAFFNQSVDEDPMLEW--290 280 296 STKAY--330

Fig. 7F

SEQUENCE LISTING

```
<110> McCarthy, Sean A
     Barnes, Thomas M
     Fraser, Christopher C
     Sharp, John D
<120> NOVEL GENES ENCODING PROTEINS HAVING DIAGNOSTIC,
     PREVENTIVE, THERAPEUTIC, AND OTHER USES
<130> 210147.0024/6PC
<140> Not Yet Assigned
<141> 2000-05-24
<150> US 09/333,159
<151> 1999-06-14
<160> 79
<170> PatentIn Ver. 2.1
<210> 1
<211> 1656
<212> DNA
<213> Homo sapiens
<400> 1
gtcgacccac gcgtccgccc acgcgtccgg cccatggcgc cgcccgccgc ccgcctcgcc 60
cccggacccg agtgtttcac agccaatggt gcggattata ggggaacaca gaactggaca 180
gcactacaag gcgggaagcc atgtctgttt tggaacgaga ctttccagca tccatacaac 240
actctgaaat accccaacgg ggagggggc ctgggtgagc acaactattg cagaaatcca 300
gatggagacg tgagcccctg gtgctatgtg gcagagcacg aggatggtgt ctactggaag 360
tactgtgaga tacctgcttg ccagatgcct ggaaaccttg gctgctacaa ggatcatgga 420
aacccacctc ctctaactgg caccagtaaa acgtccaaca aactcaccat acaaacttgc 480
atcagttttt gtcggagtca gaggttcaag tttgctggga tggagtcagg ctatgcttgc 540
ttctgtggaa acaatcctga ttactggaag tacggggagg cagccagtac cgaatgcaac 600
agcgtctgct tcggggatca cacccaaccc tgtggtggcg atggcaggat catcctcttt 660
gatactctcg tgggcgcctg cggtgggaac tactcagcca tgtcttctgt ggtctattcc 720
cctgacttcc ccgacaccta tgccacgggg agggtctgct actggaccat ccgggttccg 780
ggggcctccc acatccactt cagcttcccc ctatttgaca tcagggactc ggcggacatg 840
gtggagette tggatggeta cacccaccgt gtectagece gettecacgg gaggageege 900
ccacctctgt ccttcaacgt ctctctggac ttcgtcatct tgtatttctt ctctgatcgc 960
atcaatcagg cccagggatt tgctgtttta taccaagccg tcaaggaaga actgccacag 1020
gagaggcccg ctgtcaacca gacggtggcc gaggtgatca cggagcaggc caacctcagt 1080
gtcagcgctg eccggteete caaagteete tatgteatea ceaccageee cagecaceca 1140
```

cctcagactg tcccaggtag caattcctgg gcgccaccca tgggggctgg aagccacaga 1200

```
gttgaaggat ggacagtcta tggtctggca actctcctca tcctcacagt cacagccatt 1260
gtagcaaaga tacttctgca cgtcacattc aaatcccatc gtgttcctgc ttcaggggac 1320
cttagggatt gtcatcaacc agggacttcg ggggaaatct ggagcatttt ttacaagcct 1380
tccacttcaa tttccatctt taagaagaaa ctcaagggtc agagtcaaca agatgaccgc 1440
aatccccttg tgagtgacta aaaaccccac tgtgcctagg acttgaggtc cctctttgag 1500
ctcaaqqctq ccgtggtcaa cctctcctgt ggttcttctc tgacagactc ttccctcctc 1560
tecetetgee teggeetett eggggaaace etecteetae agaetaggaa gaggeacetg 1620
                                                                  1656
ctgccagggc aggcagagcc tggattcctc ctgctt
<210> 2
<211> 1425
<212> DNA
<213> Homo sapiens
<400> 2
atggcgccgc ccgccgcccg cctcgccctg ctctccgccg cggcgctcac gctggcggcc 60
cggcccgcgc ctagccccgg cctcggcccc ggacccgagt gtttcacagc caatggtgcg 120
gattataggg gaacacagaa ctggacagca ctacaaggcg ggaagccatg tctgttttgg 180
aacqaqactt tccaqcatcc atacaacact ctgaaatacc ccaacgggga ggggggcctg 240
ggtgagcaca actattgcag aaatccagat ggagacgtga gcccctggtg ctatgtggca 300
gagcacgagg atggtgtcta ctggaagtac tgtgagatac ctgcttgcca gatgcctgga 360
aaccttggct gctacaagga tcatggaaac ccacctcctc taactggcac cagtaaaacg 420
tccaacaaac tcaccataca aacttgcatc agtttttgtc ggagtcagag gttcaagttt 480
gctgggatgg agtcaggcta tgcttgcttc tgtggaaaca atcctgatta ctggaagtac 540
ggggaggcag ccagtaccga atgcaacagc gtctgcttcg gggatcacac ccaaccctgt 600
qqtqqcqatq qcagqatcat cctctttgat actctcqtqq qcqcctqcqq tqqqaactac 660
tragcratgt cttctgtggt ctattcccct gacttccccg acacctatgc cacggggagg 720
gtctgctact ggaccatecg ggttccgggg gcctcccaca tccacttcag cttcccccta 780
tttqacatca qqqactcqqc qqacatqqtg gagcttctgg atggctacac ccaccgtgtc 840
ctageceget tecaegggag gagecgeeca cetetgteet teaaegtete tetggaette 900
gtcatcttgt atttcttctc tgatcgcatc aatcaggccc agggatttgc tgttttatac 960
caaqccqtca aqqaaqaact qccacaqgag aggcccqctg tcaaccagac ggtggccgag 1020
gtgatcacgg agcaggccaa cctcagtgtc agcgctgccc ggtcctccaa agtcctctat 1080
gtcatcacca ccagccccag ccacccacct cagactgtcc caggtagcaa ttcctgggcg 1140
ccacccatgg gggctggaag ccacagagtt gaaggatgga cagtctatgg tctggcaact 1200
ctcctcatcc tcacagtcac agccattgta gcaaagatac ttctgcacgt cacattcaaa 1260
teccateqtq tteetgette aggggaeett agggattgte ateaaceagg gaettegggg 1320
gaaatctgga gcatttttta caagccttcc acttcaattt ccatctttaa gaagaaactc 1380
                                                                  1425
aagggtcaga gtcaacaaga tgaccgcaat ccccttgtga gtgac
<210> 3
<211> 475
```

<400> 3

<212> PRT

<213> Homo sapiens

| Met 1 | Ala | Pro | Pro | Ala 5 | Ala | Arg | Leu | Ala | Leu 10 | Leu | Ser | Ala | Ala | Ala 15 | Leu |
|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|-------------------|------------|------------|------------|
| Thr | Leu | Ala | Ala 20 | Arg | Pro | Ala | Pro | Ser 25 | Pro | Gly | Leu | Gly | Pro 30 | Gly | Pró |
| Glu | Cys | Phe 35 | Thr | Ala | Asn | Gly | Ala 40 | Asp | Tyr | Arg | Gly | Thr 45 | Gln | Asn | Trp |
| Thr | Ala 50 | Leu | Gln | Gly | Gly | Lys 55 | Pro | Cys | Leu | Phe | Trp 60 | Asn | Glu | Thr | Phe |
| Gln 65 | His | Pro | Tyr | Asn | Thr 70 | Leu | Lys | Tyr | Pro | Asn 75 | Gly | Glu | Gly | Gly | Leu 80 |
| Gly | Glu | His | Asn | Tyr 85 | Cys | Arg | Asn | Pro | Asp 90 | Gly | Asp | Val | Ser | Pro 95 | Trp |
| Суз | Tyr | Val | Ala 100 | Glu | His | Glu | Asp | Gly 105 | Val | Tyr | Trp | Lys | Туг 110 | Суз | Glu |
| Ile | Pro | Ala 115 | ĊУа | Gln | Met | Pro | Gly 120 | Asn | Leu | Gly | Суз | Tyr 125 | Lys | Asp | His |
| Gly | Asn 130 | Pro | Pro | Pro | Leu | Thr 135 | Gly | Thr | Ser | Lys | Thr 140 | Ser | Asn | Lys | Leu |
| Thr 145 | Ile | Gln | Thr | Cys | Ile 150 | Ser | Phe | Суз | Arg | Ser 155 | Gln | Arg | Phe | Lys | Phe 160 |
| Ala | Gly | Met | Glu | Ser 165 | Gly | Туг | Ala | Cys | Phe 170 | Суз | Gly | Asn | Asn | Pro 175 | Asp |
| Tyr | Trp | Lys | Tyr 180 | Gly | Glu | Ala | Ala | Ser 185 | Thr | Glu | Cys | Asn | Ser 190 | Val | Cys |
| Phe | Gly | Asp 195 | His | Thr | Gln | Pro | Суs 200 | Gly | Gly | Asp | Gly | Arg 205 | Ile | Ile | Leu |
| Phe | Asp 210 | Thr | Leu | Val | Gly | Ala 215 | Суѕ | Gly | Gly | Asn | Tyr 220 | Ser | Ala | Met | Ser |
| Ser 225 | Val | Val | Tyr | Ser | Pro 230 | Asp | Phe | Pro | Asp | Thr 235 | Tyr | Ala | Thr | Gly | Arg 240 |
| Val | Cys | Tyr | Trp | Thr 245 | Ile | Arg | Val | Pro | Gly 250 | Ala | Ser | His | Ile | His 255 | Phe |

| Ser | Phe | Pro | Leu 260 | Phe | qaA | Ile | Arg | Asp 265 | Ser | Ala | ĄsĄ | Met | Val 270 | Glu | Leu |
|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|
| Leu | Asp | Gly 275 | Tyr | Thr | His | Arg | Val 280 | Leu | Ala | Arg | Phe | His 285 | Gly | Arg | Ser |
| Arg | Pro 290 | Pro | Leu | Ser | Phe | Asn 295 | Val | Ser | Leu | Asp | Phe 300 | Val | Ile | Leu | Tyr |
| Phe 305 | Phe | Ser | Asp | Arg | Ile 310 | Asn | Gln | Ala | Gln | Gly 315 | Phe | Ala | Val | Leu | Туг 320 |
| Gln | Ala | Val | Lys | Glu 325 | Glu | Leu | Pro | Gln | Glu 330 | Arg | Pro | Ala | Val | Asn 335 | Gln |
| Thr | Val | Ala | Glu 340 | Val | Ile | Thr | Glu | Gln 345 | Ala | Asn | Leu | Ser | Val 350 | Ser | Ala |
| Ala | Arg | Ser 355 | Ser | Lys | Val | Leu | Туг 360 | Val | Ile | Thr | Thr | Ser 365 | Pro | Ser | His |
| Pro | Pro 370 | Gln | Thr | Val | Pro | Gly 375 | Ser | Asn | Ser | Trp | Ala 380 | Pro | Pro | Met | Gly |
| Ala 385 | Gly | Ser | His | Arg | Val 390 | Glu | Gly | Trp | Thr | Val 395 | Tyr | Gly | Leu | Ala | Thr 400 |
| Leu | Leu | Ile | Leu | Thr 405 | Val | Thr | Ala | Ile | Val 410 | Ala | Lys | Ile | Leu | Leu 415 | His |
| Val | Thr | Phe | Lys 420 | Ser | His | Arg | Val | Pro 425 | Ala | Ser | Gly | Asp | Leu 430 | Arg | Asp |
| Cys | His | Gln 435 | Pro | Gly | Thr | Ser | Gly 440 | Glu | Ile | Trp | Ser | Ile 445 | Phe | Tyr | Гуз |
| Pro | Ser 450 | Thr | Ser | Ile | Ser | Ile 455 | Phe | Lys | Lys | Lys | Leu 460 | Lys | Gly | Gln | Ser |
| Gln 465 | Gln | Asp | Asp | Arg | Asn 470 | Pro | Leu | Val | Ser | Asp 475 | | | | | |

<210> 4

<211> 19

<212> PRT

<213> Homo sapiens

<400> 4

Met Ala Pro Pro Ala Ala Arg Leu Ala Leu Leu Ser Ala Ala Ala Leu

1 5 10 15

Thr Leu Ala

<210> 5

<211> 456 <212> PRT <213> Homo sapiens

Thr Ala Asn Gly Ala Asp Tyr Arg Gly Thr Gln Asn Trp Thr Ala Leu 20 25 30

Gln Gly Gly Lys Pro Cys Leu Phe Trp Asn Glu Thr Phe Gln His Pro 35 40 45

Tyr Asn Thr Leu Lys Tyr Pro Asn Gly Glu Gly Gly Leu Gly Glu His
50 55 60

Asn Tyr Cys Arg Asn Pro Asp Gly Asp Val Ser Pro Trp Cys Tyr Val 65 70 75 80

Ala Glu His Glu Asp Gly Val Tyr Trp Lys Tyr Cys Glu Ile Pro Ala 85 90 95

Cys Gln Met Pro Gly Asn Leu Gly Cys Tyr Lys Asp His Gly Asn Pro 100 105 110

Pro Pro Leu Thr Gly Thr Ser Lys Thr Ser Asn Lys Leu Thr Ile Gln
115 120 125

Thr Cys Ile Ser Phe Cys Arg Ser Gln Arg Phe Lys Phe Ala Gly Met 130 135 140

Glu Ser Gly Tyr Ala Cys Phe Cys Gly Asn Asn Pro Asp Tyr Trp Lys 145 150 155 160

Tyr Gly Glu Ala Ala Ser Thr Glu Cys Asn Ser Val Cys Phe Gly Asp 165 170 175

His Thr Gln Pro Cys Gly Gly Asp Gly Arg Ile Ile Leu Phe Asp Thr

5

PCT/US00/14858 WO 00/77239

| | | | 100 | | | | | 185 | | | | | 190 | | |
|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|
| | | | 180 | | | | | | | | | | | | |
| Leu | Val | Gly 195 | Ala | Суз | Gly | Gly | Asn 200 | Tyr | Ser | Ala | Met | Ser 205 | Ser | Val | Val |
| Tyr | Ser 210 | Pro | Asp | Phe | Pro | Asp 215 | Thr | Tyr | Ala | Thr | Gly 220 | Arg | Val | Cys | Tyr |
| Trp 225 | Thr | Ile | Arg | Val | Pro 230 | Gly | Ala | Ser | His | 11e 235 | His | Phe | Ser | Phe | Pro 240 |
| Leu | Phe | Asp | Ile | Arg 245 | Asp | Ser | Ala | Asp | Met 250 | Val | Glu | Leu | Leu | Asp 255 | Gly |
| Tyr | Thr | His | Arg 260 | Val | Leu | Ala | Arg | Phe 265 | His | Gly | Arg | Ser | Arg 270 | Pro | Pro |
| Leu | Ser | Phe 275 | Asn | Val | Ser | Leu | Asp 280 | Phe | Val | Ile | Leu | Tyr 285 | Phe | Phe | Ser |
| Asp | Arg 290 | Ile | Asn | Gln | Ala | Gln 295 | Gly | . Phe | Ala | Val | Leu 300 | Tyr | Gln | Ala | Val |
| Lys 305 | | Glu | Leu | Pro | Gln 310 | Glu | Arg | Pro | Ala | Val 315 | Asn | Gln | Thr | Val | Ala 320 |
| Glu | Val | Ile | Thr | Glu 325 | | Ala | Asn | Leu | Ser 330 | Val | Ser | Ala | Ala | Arg 335 | Ser |
| Ser | Lys | Val | Leu 340 | | Val | Ile | Thr | Thr 345 | Ser | Pro | Ser | His | Pro 350 | Pro | Gln |
| Thr | Val | Pro 355 | | Ser | Asn | Ser | Trp 360 | | Pro | Pro | Met | Gly 365 | Ala | Gly | Ser |
| His | Arg 370 | | Glu | Gly | Trp | Thr 375 | | Tyr | Gly | Leu | Ala 380 | | Leu | Leu | Ile |
| Leu 385 | | · Val | . Thr | Ala | 390 | | Ala | Lys | Ile | Leu 395 | | His | Val | Thr | Phe 400 |
| Lys | Ser | His | Arg | Val 405 | | Ala | Ser | Gly | Asp 410 | | Arg | Asp | Cys | His 415 | Gln |
| Pro | Gly | Thi | Ser 420 | | g Glu | Ile | Trp | Ser 425 | | Phe | туг | Lys | Pro 430 | | Thr |

Ser Ile Ser Ile Phe Lys Lys Lys Leu Lys Gly Gln Ser Gln Gln Asp

435 440 445

Asp Arg Asn Pro Leu Val Ser Asp 450 455

<210> 6

<211> 373

<212> PRT

<213> Homo sapiens

<400> 6

Ala Arg Pro Ala Pro Ser Pro Gly Leu Gly Pro Gly Pro Glu Cys Phe
1 5 10 15

Thr Ala Asn Gly Ala Asp Tyr Arg Gly Thr Gln Asn Trp Thr Ala Leu 20 25 30

Gln Gly Gly Lys Pro Cys Leu Phe Trp Asn Glu Thr Phe Gln His Pro 35 40 45

Tyr Asn Thr Leu Lys Tyr Pro Asn Gly Glu Gly Gly Leu Gly Glu His
50 55 60

Asn Tyr Cys Arg Asn Pro Asp Gly Asp Val Ser Pro Trp Cys Tyr Val 65 70 75 80

Ala Glu His Glu Asp Gly Val Tyr Trp Lys Tyr Cys Glu Ile Pro Ala 85 90 95

Cys Gln Met Pro Gly Asn Leu Gly Cys Tyr Lys Asp His Gly Asn Pro 100 105 110

Pro Pro Leu Thr Gly Thr Ser Lys Thr Ser Asn Lys Leu Thr Ile Gln
115 120 125

Thr Cys Ile Ser Phe Cys Arg Ser Gln Arg Phe Lys Phe Ala Gly Met 130 135 140

Glu Ser Gly Tyr Ala Cys Phe Cys Gly Asn Asn Pro Asp Tyr Trp Lys 145 150 155 160

Tyr Gly Glu Ala Ala Ser Thr Glu Cys Asn Ser Val Cys Phe Gly Asp 165 170 175

His Thr Gln Pro Cys Gly Gly Asp Gly Arg Ile Ile Leu Phe Asp Thr 180 185 190

Leu Val Gly Ala Cys Gly Gly Asn Tyr Ser Ala Met Ser Ser Val Val 195 200 Tyr Ser Pro Asp Phe Pro Asp Thr Tyr Ala Thr Gly Arg Val Cys Tyr 220 215 Trp Thr Ile Arg Val Pro Gly Ala Ser His Ile His Phe Ser Phe Pro 235 230 Leu Phe Asp Ile Arg Asp Ser Ala Asp Met Val Glu Leu Leu Asp Gly 250 245 Tyr Thr His Arg Val Leu Ala Arg Phe His Gly Arg Ser Arg Pro Pro 265 Leu Ser Phe Asn Val Ser Leu Asp Phe Val Ile Leu Tyr Phe Phe Ser 275 280 285 Asp Arg Ile Asn Gln Ala Gln Gly Phe Ala Val Leu Tyr Gln Ala Val 295 300 290 Lys Glu Glu Leu Pro Gln Glu Arg Pro Ala Val Asn Gln Thr Val Ala 305 310 315 Glu Val Ile Thr Glu Gln Ala Asn Leu Ser Val Ser Ala Ala Arg Ser 325 330 Ser Lys Val Leu Tyr Val Ile Thr Thr Ser Pro Ser His Pro Pro Gln 340 345 Thr Val Pro Gly Ser Asn Ser Trp Ala Pro Pro Met Gly Ala Gly Ser 360 His Arg Val Glu Gly 370 <210> 7 <211> 23 <212> PRT <213> Homo sapiens <400> 7

Ile Val Ala Lys Ile Leu Leu

20

5

Trp Thr Val Tyr Gly Leu Ala Thr Leu Leu Ile Leu Thr Val Thr Ala

10

```
<210> 8
<211> 60
<212> PRT
<213> Homo sapiens
<400> 8
His Val Thr Phe Lys Ser His Arg Val Pro Ala Ser Gly Asp Leu Arg
 1
                 5
                                   10
Asp Cys His Gln Pro Gly Thr Ser Gly Glu Ile Trp Ser Ile Phe Tyr
                               25
Lys Pro Ser Thr Ser Ile Ser Ile Phe Lys Lys Leu Lys Gly Gln
                            40
Ser Gln Gln Asp Asp Arg Asn Pro Leu Val Ser Asp
                        55
<210> 9
<211> 4628
<212> DNA
<213> Homo sapiens
<400> 9
gcggccgctc gcgatctaga actagtaatg atgctgcctc aaaactcgtg gcatattgat 60
tttggaagat getgetgtea teagaacett ttetetgetg tggtaacetg cateetgete 120
ctgaattcct gctttctcat cagcagtttt aatggaacag atttggagtt gaggctggtc 180
aatggagacg gtccctgctc tgggacagtg gaggtgaaat tccagggaca gtgggggact 240
gtgtgtgatg atgggtggaa cactactgcc tcaactgtcg tgtgcaaaca gcttggatgt 300
ccattttctt tcgccatgtt tcgttttgga caagccgtga ctagacatgg aaaaatttgg 360
cttgatgatg tttcctgtta tggaaatgag tcagctctct gggaatgtca acaccgggaa 420
tggggaagcc ataactgtta tcatggagaa gatgttggtg tgaactgtta tggtgaagcc 480
aatctgggtt tgaggctagt ggatggaaac aactcctgtt cagggagagt ggaqgtgaaa 540
ttccaagaaa ggtgggggac tatatgtgat gatgggtgga acttgaatac tgctgccgtg 600
gtgtgcaggc aactaggatg tccatcttct tttatttctt ctggagttgt taatagccct 660
gctgtattgc gccccatttg gctggatgac attttatgcc aggggaatga gttggcactc 720
tggaattgca gacatcgtgg atggggaaat catgactgca gtcacaatga ggatgtcaca 780
ttaacttgtt atgatagtag tgatcttgaa ctaaggcttg taggtggaac taaccgctgt 840
atggggagag tagagctgaa aatccaagga aggtggggga ccgtatgcca ccataagtgg 900
gctggcttgc ctcatttgca gtcagggtct gatqttgtat ggcttgatgg tqtctcctqc 1020
teeggtaatg aatetttet ttgggaetge agacatteeg gaacegteaa ttttgaetgt 1080
cttcatcaaa acgatgtgtc tgtgatctgc tcagatggag cagatttgga actqcqacta 1140
gcaqatqqaa qtaacaattq ttcaqgqaqa qtaqaqqtqa qaattcatga acaqtqqtqq 1200
acaatatgtg accagaactg gaagaatgaa caagcccttg tggtttgtaa gcagctagga 1260
```

9

| tgtccgttca | gcgtctttgg | cagtcgtcgt | gctaaaccta | gtaatgaagc | tagagacatt | 1320 |
|-------------|------------|------------|------------|------------|------------|------|
| | gcatatcttg | | | | | |
| | agcgaacatg | | | | • | |
| | acctaaggct | | | | | |
| | gagagtgggg | | | | | |
| | aacaattggg | | | | | |
| | caggacctat | | | | | |
| | gtgaacacag | | | | | |
| | gctcaggtga | | | | | |
| | gactggaggt | | | | | |
| | aagctgcagc | | | | | |
| ggcatgggtc | tgggaaacgc | ttctacagga | tatggaaaaa | tttggctcga | tgatgtttcc | 1980 |
| tgtgatggag | atgagtcaga | tctctggtca | tgcaggaaca | gtgggtgggg | aaataatgac | 2040 |
| tgcagtcaca | gtgaagatgt | tggagtgatc | tgttctgatg | catcggatat | ggagctgagg | 2100 |
| cttgtgggtg | gaagcagcag | gtgtgctgga | aaagttgagg | tgaatgtcca | gggtgccgtg | 2160 |
| ggaattctgt | gtgctaatgg | ctggggaatg | aacattgctg | aagttgtttg | caggcaactt | 2220 |
| gaatgtgggt | ctgcaatcag | ggtctccaga | gagcctcatt | tcacagaaag | aacattacac | 2280 |
| atcttaatgt | cgaattctgg | ctgcactgga | ggggaagcct | ctctctggga | ttgtatacga | 2340 |
| tgggagtgga | aacagactgc | gtgtcattta | aatatggaag | caagtttgat | ctgctcagcc | 2400 |
| cacaggcagc | ccaggctggt | tggagctgat | atgccctgct | ctggacgtgt | tgaagtgaaa | 2460 |
| catgcagaca | catggcgctc | tgtctgtgat | tctgatttct | ctcttcatgc | tgccaatgtg | 2520 |
| ctgtgcagag | aattaaattg | tggagatgcc | atatctcttt | ctgtgggaga | tcactttgga | 2580 |
| aaagggaatg | gtctaacttg | ggccgaaaag | ttccagtgtg | aagggagtga | aactcacctt | 2640 |
| gcattatgcc | ccattgttca | acatccggaa | gacacttgta | tccacagcag | agaagttgga | 2700 |
| gttgtctgtt | cccgatatac | agatgtccga | cttgtgaatg | gcaaatccca | gtgtgacggg | 2760 |
| caagtggaga | tcaacgtgct | tggacactgg | ggctcactgt | gtgacaccca | ctgggaccca | 2820 |
| gaagatgccc | gtgttctatg | cagacagete | agctgtggga | ctgctctctc | aaccacagga | 2880 |
| ggaaaatata | ttggagaaag | aagtgttcgt | gtgtggggac | acaggtttca | ttgcttaggg | 2940 |
| aatgagtcac | ttctggataa | ctgtcaaatg | acagttcttg | gagcacctcc | ctgtatccat | 3000 |
| ggaaatactg | tctctgtgat | ctgcacagga | agcctgaccc | agccactgtt | tccatgcctc | 3060 |
| gcaaatgtat | ctgacccata | tttgtctgca | gttccagagg | gcagtgcttt | gatctgctta | 3120 |
| gaggacaaac | ggctccgcct | agtggatggg | gacagccgct | gtgccgggag | agtagagatc | 3180 |
| tatcacgacg. | gcttctgggg | caccatctgt | gatgacggct | gggacctgag | cgatgcccac | 3240 |
| gtggtgtgtc | aaaagctggg | ctgtggagtg | gccttcaatg | ccacggtctc | tgctcacttt | 3300 |
| ggggaggggt | cagggcccat | ctggctggat | gacctgaact | gcacaggaac | ggagtcccac | 3360 |
| ttgtggcagt | gcccttcccg | cggctggggg | cagcacgact | gcaggcacaa | ggaggacgca | 3420 |
| ggggtcatct | gctcagaatt | cacagccttg | aggctctaca | gtgaaactga | aacagagagc | 3480 |
| tgtgctggga | gattggaagt | cttctataac | gggacctggg | gcagcgtcgg | caggaggaac | 3540 |
| atcaccacag | ccatagcagg | cattgtgtgc | aggcagctgg | gctgtgggga | gaatggagtt | 3600 |
| gtcagcctcg | cccctttatc | taagacaggc | tctggtttca | tgtgggtgga | tgacattcag | 3660 |
| | cgcatatete | | | | | |
| | cagaagagac | | | | | |
| | gctctgggag | | | | | |
| | gggacctggc | | | | | |
| | ccctgaggga | | | | | |
| | gcaaaggaaa | | | | | |
| | gtggacacaa | | | | | |
| | cctcctcagg | | | | | |
| _ | | _ | | | | |

```
ctcctggttc tgtttattct atttctacg tggtgccgag ttcagaaaca aaaacatctg 4200 cccctcagag tttcaaccag aaggagggt tctctcgagg agaatttatt ccatgagatg 4260 gagacctgcc tcaagagag ggacccacat gggacaagaa cctcagatga cacccccaac 4320 catggttgtg aagatgctag cgacacatcg ctgttgggag ttcttcctgc ctctgaagcc 4380 acaacaacta ttaaatgaat aaagaggaag tcaagttgcc ctatggaaaa cttgaaggag 4440 acaacattct tgaacaatag gagaacagct aaattgataa agactggtga taataaaaa 4560 tgaattatg atatcactgt taaaaaaaaa aaaaaaaaa aaaaaaaaa acggacggt 4620 gggtcgac
```

<210> 10 <211> 4359 <212> DNA <213> Homo sapiens

<400> 10

```
atgatgctgc ctcaaaactc gtggcatatt gattttggaa gatgctgctg tcatcagaac 60
cttttctctg ctgtggtaac ttgcatcctg ctcctgaatt cctgctttct catcagcagt 120
tttaatggaa cagatttgga gttgaggctg gtcaatggag acggtccctg ctctgggaca 180
qtqqaqqtqa aattccaqqq acaqtggggq actqtgtgtg atgatgggtg gaacactact 240
gcctcaactg tcgtgtgcaa acagcttgga tgtccatttt ctttcgccat gtttcgtttt 300
ggacaagccg tgactagaca tggaaaaatt tggcttgatg atgtttcctg ttatggaaat 360
gagtcagctc tctgggaatg tcaacaccgg gaatggggaa gccataactg ttatcatgga 420
gaagatgttg gtgtgaactg ttatggtgaa gccaatctgg gtttgaggct agtggatgga 480
aacaactcct gttcagggag agtggaggtg aaattccaag aaaggtgggg gactatatgt 540
gatgatgggt ggaacttgaa tactgctgcc gtggtgtgca ggcaactagg atgtccatct 600
tettttattt ettetggagt tgttaatage eetgetgtat tgegeeecat ttggetggat 660
gacattttat gccaggggaa tgagttggca ctctggaatt gcagacatcg tggatgggga 720
aatcatgact gcagtcacaa tgaggatgtc acattaactt gttatgatag tagtgatctt 780
gaactaaggc ttgtaggtgg aactaaccgc tgtatgggga gagtagagct gaaaatccaa 840
ggaaggtggg ggaccgtatg ccaccataag tggaacaatg ctgcagctga tgtcgtatgc 900
aagcagttgg gatgtggaac cgcacttcac ttcgctggct tgcctcattt gcagtcaggg 960
tetgatgttg tatggettga tggtgtetee tgeteeggta atgaatettt tetttgggae 1020
tgcagacatt ccggaaccgt caattttgac tgtcttcatc aaaacgatgt gtctgtgatc 1080
tgctcagatg gagcagattt ggaactgcga ctagcagatg gaagtaacaa ttgttcaggg 1140
agagtagagg tgagaattca tgaacagtgg tggacaatat gtgaccagaa ctggaagaat 1200
gaacaagccc ttgtggtttg taagcagcta ggatgtccgt tcagcgtctt tggcagtcgt 1260
cgtgctaaac ctagtaatga agctagagac atttggataa acagcatatc ttgcactggg 1320
aatgagtcag ctctctggga ctgcacatat gatggaaaag caaagcgaac atgcttccga 1380
agatcagatg ctggagtaat ttgttctgat aaggcagatc tggacctaag gcttgtcggg 1440
gctcatagcc cctgttatgg gagattggag gtgaaatacc aaggagagtg ggggactgtg 1500
tgtcatgaca gatggagcac aaggaatgca gctgttgtgt gtaaacaatt gggatgtgga 1560
aagcetatge atgtgtttgg tatgacetat tttaaagaag catcaggace tatttggetg 1620
gatgacgttt cttgcattgg aaatgagtca aatatctggg actgtgaaca cagtggatgg 1680
ggaaaqcata attqtqtaca cagaqaqgat gtgattqtaa cctgctcagg tgatqcaaca 1740
tggggcctga ggctqgtqgg cggcagcaac cgctgctcgg gaagactgga ggtgtacttt 1800
caaggacggt ggggcacagt gtgtgatgac ggctggaaca gtaaagctgc agctgtggtg 1860
```

11

```
tgtagccagc tggactgccc atcttctatc attggcatgg gtctgggaaa cgcttctaca 1920
ggatatggaa aaatttggct cgatgatgtt tcctgtgatg gagatgagtc agatctctgg 1980
tcatgcagga acagtgggtg gggaaataat gactgcagtc acagtgaaga tgttggagtg 2040
atctgttctg atgcatcgga tatggagctg aggcttgtgg gtggaagcag caggtgtgct 2100
ggaaaagttg aggtgaatgt ccagggtgcc gtgggaattc tgtgtgctaa tggctgggga 2160
atgaacattg ctgaagttgt ttgcaggcaa cttgaatgtg ggtctgcaat cagggtctcc 2220
agagageete attteacaga aagaacatta cacatettaa tgtegaatte tggetgeaet 2280
ggaggggaag cetetetetg ggattgtata egatgggagt ggaaacagae tgegtgteat 2340
ttaaatatgg aagcaagttt gatctgctca gcccacaggc agcccaggct ggttggagct 2400
gatatgeeet getetggaeg tgttgaagtg aaacatgeag acacatggeg etetgtetgt 2460
gattctgatt tctctcttca tgctgccaat gtgctgtgca gagaattaaa ttgtggagat 2520
gccatatctc tttctgtggg agatcacttt ggaaaaggga atggtctaac ttgggccgaa 2580
aagttccagt gtgaagggag tgaaactcac cttgcattat gccccattgt tcaacatccg 2640
gaagacactt gtatccacag cagagaagtt ggagttgtct gttcccgata tacagatgtc 2700
cgacttgtga atggcaaatc ccagtgtgac gggcaagtgg agatcaacgt gcttggacac 2760
tggggctcac tgtgtgacac ccactgggac ccagaagatg cccgtgttct atgcagacag 2820
ctcagctgtg ggactgctct ctcaaccaca ggaggaaaat atattggaga aagaagtgtt 2880
cgtgtgtggg gacacaggtt tcattgctta gggaatgagt cacttctgga taactgtcaa 2940
atgacagttc ttggagcacc tccctgtatc catggaaata ctgtctctgt gatctgcaca 3000
ggaageetga eccageeact gtttecatge etegeaaatg tatetgaeee atatttgtet 3060
gcagttccag agggcagtgc tttgatctgc ttagaggaca aacggctccg cctagtggat 3120
ggggacagec getgtgeegg gagagtagag atetateaeg aeggettetg gggcaccate 3180
tgtgatgacg gctgggacct gagcgatgcc cacgtggtgt gtcaaaagct gggctgtgga 3240
gtggccttca atgccacggt ctctgctcac tttggggagg ggtcagggcc catctggctg 3300
gatgacetga actgcacagg aacggagtee caettgtgge agtgceette eegeggetgg 3360
gggcagcacg actgcaggca caaggaggac gcaggggtca tctgctcaga attcacagcc 3420
ttgaggctct acagtgaaac tgaaacagag agctgtgctg ggagattgga agtcttctat 3480
aacgggacct ggggcagcgt cggcaggagg aacatcacca cagccatagc aggcattgtg 3540
tgcaggcagc tgggctgtgg ggagaatgga gttgtcagcc tcgccccttt atctaagaca 3600
ggctctggtt tcatgtgggt ggatgacatt cagtgtccta aaacgcatat ctccatatgg 3660
cagtgcctgt ctgccccatg ggagcgaaga atctccagcc cagcagaaga gacctggatc 3720
acatgtgaag atagaataag agtgcgtgga ggagacaccg agtgctctgg gagagtggag 3780
atctggcacg caggctcctg gggcacagtg tgtgatgact cctgggacct ggccgaggcg 3840
gaagtggtgt gtcagcagct gggctgtggc tctgctctgg ctgccctgag ggacgcttcg 3900
tttggccagg gaactggaac catctggttg gatgacatgc ggtgcaaagg aaatgagtca 3960
tttctatggg actgtcacgc caaaccctgg ggacagagtg actgtggaca caaggaagat 4020
gctggcgtga ggtgctctgg acagtcgctg aaatcactga atgcctcctc aggtcattta 4080
gcacttattt tatccagtat ctttgggctc cttctcctgg ttctgtttat tctatttctc 4140
acgtggtgcc gagttcagaa acaaaaacat ctgcccctca gagtttcaac cagaaggagg 4200
qqttctctcg aggagaattt attccatgag atggagacct gcctcaagag agaggaccca 4260
catgggacaa gaacctcaga tgacaccccc aaccatggtt gtgaagatgc tagcgacaca 4320
tegetgttgg gagttettee tgeetetgaa geeacaaaa
                                                                   4359
```

<210> 11 <211> 1453 <212> PRT

<213> Homo sapiens

| | > 11 | | | | | | | | | | | | | | |
|------------|------------|------------|------------|------------|--------------|------------|------------|------------|------------|------------|------------|------------|------------|-----------|------------|
| Met 1 | Met | Leu | Pro | Gln 5 | Asn : | Ser | Trp | His | lle 10 | Asp | Phe | Gly | Arg | Cys 15 | Суз |
| Cys | His | Gln | Asn 20 | Leu | Phe | Ser | Ala | Val 25 | Val | Thr | Cys | Ile | Leu 30 | Leu | Leu |
| Asn | Ser | Cys 35 | Phe | Leu | Ile | Ser | Ser 40 | Phe | Asn | Gly | Thr | Asp 45 | Leu | Glu | Leu |
| Arg | Leu 50 | Val | Asn | Gly | Asp | Gly 55 | Pro | Cys | Ser | Gly | Thr 60 | Val | Glu | Val | Lys |
| Phe 65 | Gln | Gly | Gln | Trp | Gly 70 | Thr | Val | Суз | Asp | Asp 75 | Gly | Trp | Asn | Thr | Thr 80 |
| Ala | Ser | Thr | Val | Val 85 | Суѕ | Lys | Gln | Leu | Gly 90 | Суз | Pro | Phe | Ser | Phe 95 | Ala |
| Met | Phe | Arg | Phe 100 | Gly | Gln | Ala | Val | Thr 105 | Arg | His | Gly | Lys | 11e | Trp | Leu |
| Asp | Asp | Val 115 | Ser | Cys | Туг | Gly | Asn 120 | Glu | Ser | Ala | Leu | Trp 125 | Glu | Cys | Gln |
| His | Arg 130 | | Trp | Gly | Ser | His 135 | Asn | | Tyr | His | Gly 140 | Glu | Asp | Val | Gly |
| Val 145 | | Cys | Tyr | Gly | Glu 150 | Ala | Asn | Leu | Gly | Leu 155 | | Leu | Val | Asp | Gly 160 |
| Asn | Asn | Ser | Cys | Ser 165 | | Arg | Val | Glu | Val 170 | | Phe | Gln | Glu | 175 | Trp |
| Gly | Thr | Ile | Cys | | Asp | Gly | Trp | Asn 185 | Leu | neA | Thr | Ala | Ala 190 | Val | Val |
| Суs | Arg | Gln 195 | | ı Gly | у Суз | Pro | Ser 200 | | Pĥe | : Ile | Ser | Ser 205 | | v Val | . Val |
| Asr | Sez 210 | | Ala | a Val | Leu | Arg 215 | | Ile | Trp | Leu | 220 | | ıle | e Leu | ı Cys |
| Glr 225 | | y Asr | ı Glı | ı Let | a Ala 230 | | Trp | Asn | Cys | 235 | | Arc | g Gl | / Tr | Gly 240 |
| Asr | n His | a Ası | Cy: | s Sei | c His | . Asn | Glu | Asp | Va] | L Thr | Leu | Thi | c Cys | з Ту | r Asp |

PCT/US00/14858 WO 00/77239

| | | | | 245 | | | | | 250 | | | | | 255 | |
|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|
| Ser | Ser | Asp | Leu 260 | Glu | Leu | Arg | Leu | Val 265 | Gly | Gly | Thr | Asn | Arg 270 | Cys | Met |
| Gly | Arg | Val 275 | Glu | Leu | Lys | Ile | Gln 280 | Gly | Arg | Trp | Gly | Thr 285 | Val | Cys | His |
| His | Lys 290 | Trp | Asn | Asn | Ala | Ala 295 | Ala | Asp | Val | Val | Cys 300 | Lys | Gln | Leu | Gly |
| Cys 305 | Gly | Thr | Ala | Leu | His 310 | Phe | Ala | Gly | Leu | Pro 315 | His | Leu | Gln | Ser | Gly 320 |
| Ser | Asp | Val | Val | Trp 325 | Leu | ĄżĄ | Gly | Val | Ser 330 | Суз | Ser | Gly | Asn | Glu 335 | Ser |
| Phe | Leu | Trp | Asp 340 | Суз | Arg | His | Ser | Gly 345 | Thr | Val | Asn | Phe | Asp 350 | Суз | Leu |
| His | Gln | Asn 355 | Asp | Val | Ser | Val | 11e 360 | Суз | Ser | Asp | Gly | Ala 365 | Asp | Leu | Glu |
| Leu | Arg 370 | Leu | Ala | Asp | Gly | Ser 375 | Asn | Asn | Cys | Ser | Gly 380 | Arg | Val | Glu | Val |
| Arg 385 | Ile | His | Glu | Gln | Trp 390 | Trp | Thr | Ile | Суз | Asp 395 | Gln | Asn | Trp | Lys | Asn 400 |
| Glu | Gln | Ala | Leu | Val 405 | Val | Суз | Lys | Gln | Leu 410 | Gly | Cys | Pro | Phe | Ser 415 | Val |
| Phe | Gly | Ser | Arg 420 | Arg | Ala | Lys | Pro | Ser 425 | Asn | Glu | Ala | Arg | Asp 430 | Ile | Trp |
| Ile | Asn | Ser 435 | Ile | Ser | Cys | Thr | Gly 440 | Asn | Glu | Ser | Ala | Leu 445 | Trp | Asp | Суз |
| Thr | Tyr 450 | Asp | Gly | Lys | Ala | Lys 455 | Arg | Thr | Cys | Phe | Arg 460 | Arg | Ser | Asp | Ala |
| Gly 465 | Val | Ile | Суз | Ser | Asp 470 | Lys | Ala | Asp | Leu | Asp 475 | Leu | Arg | Leu | Val | Gly 480 |
| Ala | His | Ser | Pro | Cys 485 | Tyr | Gly | Arg | Leu | Glu 490 | Val | ГÀЗ | Tyr | Gln | Gly 495 | Glu |

Trp Gly Thr Val Cys His Asp Arg Trp Ser Thr Arg Asn Ala Ala Val

WO 00/77239

Val Cys Lys Gln Leu Gly Cys Gly Lys Pro Met His Val Phe Gly Met Thr Tyr Phe Lys Glu Ala Ser Gly Pro Ile Trp Leu Asp Asp Val Ser Cys Ile Gly Asn Glu Ser Asn Ile Trp Asp Cys Glu His Ser Gly Trp Gly Lys His Asn Cys Val His Arg Glu Asp Val Ile Val Thr Cys Ser Gly Asp Ala Thr Trp Gly Leu Arg Leu Val Gly Gly Ser Asn Arg Cys Ser Gly Arg Leu Glu Val Tyr Phe Gln Gly Arg Trp Gly Thr Val Cys Asp Asp Gly Trp Asn Ser Lys Ala Ala Ala Val Val Cys Ser Gln Leu Asp Cys Pro Ser Ser Ile Ile Gly Met Gly Leu Gly Asn Ala Ser Thr Gly Tyr Gly Lys Ile Trp Leu Asp Asp Val Ser Cys Asp Gly Asp Glu Ser Asp Leu Trp Ser Cys Arg Asn Ser Gly Trp Gly Asn Asn Asp Cys Ser His Ser Glu Asp Val Gly Val Ile Cys Ser Asp Ala Ser Asp Met Glu Leu Arg Leu Val Gly Gly Ser Ser Arg Cys Ala Gly Lys Val Glu Val Asn Val Gln Gly Ala Val Gly Ile Leu Cys Ala Asn Gly Trp Gly Met Asn Ile Ala Glu Val Val Cys Arg Gln Leu Glu Cys Gly Ser Ala Ile Arg Val Ser Arg Glu Pro His Phe Thr Glu Arg Thr Leu His Ile

Leu Met Ser Asn Ser Gly Cys Thr Gly Glu Ala Ser Leu Trp Asp

| | | 755 | | | | | 760 | | | | | 765 | | | |
|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|
| Cys | Ile 770 | Arg | Trp | Glu | Trp | Lys 775 | Gln | Thr | Ala | Cys | His 780 | Leu | Asn | Met | Glu |
| Ala 785 | Ser | Leu | Ile | Суѕ | Ser 790 | Ala | His | Arg | Gln | Pro 795 | Arg | Leu | Val | Gly | Ala 800 |
| Asp | Met | Pro | Cys | Ser 805 | Gly | Arg | Val | Glu | Val 810 | Lys | His | Ala | Asp | Thr 815 | Trp |
| Arg | Ser | Val | Суз 820 | Asp | Ser | Asp | Phe | Ser 825 | Leu | His | Ala | Ala | Asn 830 | Val | Leu |
| Cys | Arg | Glu 835 | Leu | Asn | Cys | Gly | Asp 840 | Ala | Ile | Ser | Гел | Ser 845 | Val | Gly | Asp |
| His | Phe 850 | Gly | Lys | Gly | Asn | Gly 855 | Leu | Thr | Trp | Ala | Glu 860 | Lys | Phe | Gln | Суз |
| Glu 865 | Gly | Ser | Glu | Thr | His 870 | Leu | Ala | Leu | Суз | Pro 875 | Ile | Val | Gln | His | Pro 880 |
| Glu | Asp | Thr | Суѕ | 11e 885 | His | Ser | Arg | Glu | Val 890 | Gly | Val | Val | Суз | Ser 895 | Arg |
| Tyr | Thr | Asp | Val 900 | Arg | Leu | Val | Asn | Gly 905 | Lys | Ser | Gln | Суз | Asp 910 | Gly | Gln |
| Val | Glu | Ile 915 | Asn | Val | Leu | Gly | His 920 | Trp | Gly | Ser | Leu | Суs 925 | Asp | Thr | His |
| Trp | Asp 930 | Pro | Glu | Asp | Ala | Arg 935 | Val | Leu | Cys | Arg | Gln 940 | Leu | Ser | Суз | Gly |
| Thr 945 | Ala | Leu | Ser | Thr | Thr 950 | Gly | Gly | Lys | Tyr | Ile 955 | Gly | Glu | Arg | Ser | Val 960 |
| Arg | Val | Trp | Gly | His 965 | Arg | Phe | His | Cys | Leu 970 | Gly | Asn | Glu | Ser | Leu 975 | Leu |
| Asp | Asn | Cys | Gln 980 | Met | Thr | Val | Leu | Gly 985 | Ala | Pro | Pro | Cys | Ile 990 | His | Gly |
| Asn | Thr | Val | Ser | Val | Ile | _ | Thr | Gly | Ser | Leu | Thr | Gln | Pro | Leu | Phe |

Pro Cys Leu Ala Asn Val Ser Asp Pro Tyr Leu Ser Ala Val Pro Glu

1010 1015 1020

Gly Ser Ala Leu Ile Cys Leu Glu Asp Lys Arg Leu Arg Leu Val Asp 1025 1030 1035 1040

- Gly Asp Ser Arg Cys Ala Gly Arg Val Glu Ile Tyr His Asp Gly Phe 1045 1050 1055
- Trp Gly Thr Ile Cys Asp Asp Gly Trp Asp Leu Ser Asp Ala His Val 1060 1065 1070
- Val Cys Gln Lys Leu Gly Cys Gly Val Ala Phe Asn Ala Thr Val Ser 1075 1080 1085
- Ala His Phe Gly Glu Gly Ser Gly Pro Ile Trp Leu Asp Asp Leu Asn 1090 1095 1100
- Cys Thr Gly Thr Glu Ser His Leu Trp Gln Cys Pro Ser Arg Gly Trp 1105 1110 1115 1120
- Gly Gln His Asp Cys Arg His Lys Glu Asp Ala Gly Val Ile Cys Ser 1125 1130 1135
- Glu Phe Thr Ala Leu Arg Leu Tyr Ser Glu Thr Glu Thr Glu Ser Cys 1140 1145 1150
- Ala Gly Arg Leu Glu Val Phe Tyr Asn Gly Thr Trp Gly Ser Val Gly
 1155 1160 1165
- Arg Arg Asn Ile Thr Thr Ala Ile Ala Gly Ile Val Cys Arg Gln Leu 1170 1175 1180
- Gly Cys Gly Glu Asn Gly Val Val Ser Leu Ala Pro Leu Ser Lys Thr 1185 1190 1195 1200
- Gly Ser Gly Phe Met Trp Val Asp Asp Ile Gln Cys Pro Lys Thr His 1205 1210 1215
- Ile Ser Ile Trp Gln Cys Leu Ser Ala Pro Trp Glu Arg Arg Ile Ser 1220 1225 1230
- Ser Pro Ala Glu Glu Thr Trp Ile Thr Cys Glu Asp Arg Ile Arg Val 1235 1240 1245
- Arg Gly Gly Asp Thr Glu Cys Ser Gly Arg Val Glu Ile Trp His Ala 1250 1255 1260
- Gly Ser Trp Gly Thr Val Cys Asp Ser Trp Asp Leu Ala Glu Ala

Glu Val Val Cys Gln Gln Leu Gly Cys Gly Ser Ala Leu Ala Ala Leu Arg Asp Ala Ser Phe Gly Gln Gly Thr Gly Thr Ile Trp Leu Asp Asp Met Arg Cys Lys Gly Asn Glu Ser Phe Leu Trp Asp Cys His Ala Lys Pro Trp Gly Gln Ser Asp Cys Gly His Lys Glu Asp Ala Gly Val Arg Cys Ser Gly Gln Ser Leu Lys Ser Leu Asn Ala Ser Ser Gly His Leu Ala Leu Ile Leu Ser Ser Ile Phe Gly Leu Leu Leu Val Leu Phe Ile Leu Phe Leu Thr Trp Cys Arg Val Gln Lys Gln Lys His Leu Pro Leu Arg Val Ser Thr Arg Arg Gly Ser Leu Glu Glu Asn Leu Phe His Glu Met Glu Thr Cys Leu Lys Arg Glu Asp Pro His Gly Thr Arg Thr Ser Asp Asp Thr Pro Asn His Gly Cys Glu Asp Ala Ser Asp Thr Ser Leu Leu Gly Val Leu Pro Ala Ser Glu Ala Thr Lys <210> 12 <211> 40 <212> PRT <213> Homo sapiens <400> 12 Met Met Leu Pro Gln Asn Ser Trp His Ile Asp Phe Gly Arg Cys Cys Cys His Gln Asn Leu Phe Ser Ala Val Val Thr Cys Ile Leu Leu

Asn Ser Cys Phe Leu Ile Ser Ser 35 40

<210> 13

<211> 1413

<212> PRT

<213> Homo sapiens

<400> 13

Phe Asn Gly Thr Asp Leu Glu Leu Arg Leu Val Asn Gly Asp Gly Pro
1 5 10 15

Cys Ser Gly Thr Val Glu Val Lys Phe Gln Gly Gln Trp Gly Thr Val
20 25 30

Cys Asp Asp Gly Trp Asn Thr Thr Ala Ser Thr Val Val Cys Lys Gln
35 40 45

Leu Gly Cys Pro Phe Ser Phe Ala Met Phe Arg Phe Gly Gln Ala Val 50 55 60

Thr Arg His Gly Lys Ile Trp Leu Asp Asp Val Ser Cys Tyr Gly Asn 65 70 75 80

Glu Ser Ala Leu Trp Glu Cys Gln His Arg Glu Trp Gly Ser His Asn 85 90 95

Cys Tyr His Gly Glu Asp Val Gly Val Asn Cys Tyr Gly Glu Ala Asn 100 105 ~110

Leu Gly Leu Arg Leu Val Asp Gly Asn Asn Ser Cys Ser Gly Arg Val 115 120 125

Glu Val Lys Phe Gln Glu Arg Trp Gly Thr Ile Cys Asp Asp Gly Trp 130 135 140

Asn Leu Asn Thr Ala Ala Val Val Cys Arg Gln Leu Gly Cys Pro Ser 145 150 155 160

Ser Phe Ile Ser Ser Gly Val Val Asn Ser Pro Ala Val Leu Arg Pro 165 170 175

Ile Trp Leu Asp Asp Ile Leu Cys Gln Gly Asn Glu Leu Ala Leu Trp 180 185 190

Asn Cys Arg His Arg Gly Trp Gly Asn His Asp Cys Ser His Asn Glu 195 200 205

| Asp | Val 210 | Thr | Leu | Thr | Суз | Tyr 215 | Asp | Ser | Ser | Asp | Leu 220 | Glu | Leu | Arg | Leu |
|------------|------------|------------|------------|------------|------------|------------|------------|------------|-------------------|------------|------------|------------|------------|------------|------------|
| Val 225 | Gly | Gly | Thr | Asn | Arg 230 | Суз | Met | Gly | Arg | Val 235 | Glu | Leu | Lys | Ile | Gln 240 |
| Gly | Arg | Trp | Gly | Thr 245 | Val | Cys | His | His | Lys 250 | Trp | Asn | Asn | Ala | Ala 255 | Ala |
| Asp | Val | Val | Cys 260 | Lys | Gln | Leu | Gly | Cys 265 | Gly | Thr | Ala | Leu | His 270 | Phe | Ala |
| Gly | Leu | Pro 275 | His | Leu | Gln | Ser | Gly 280 | Ser | Asp | Val | Val | Trp 285 | Leu | Asp | Gly |
| Val | Ser 290 | Суѕ | Ser | Gly | Asn | Glu 295 | Ser | Phe | Leu | Trp | Asp 300 | Суз | Arg | His | Ser |
| Gly 305 | Thr | Val | Asn | Phe | Asp 310 | Cys | Leu | His | Gln | Asn 315 | Asp | Val | Ser | Val | 11e 320 |
| Суз | Ser | Asp | Gly | Ala 325 | Asp | Leu | Glu | Leu | Arg 330 | Leu | Ala | Asp | Gly | Ser 335 | Asn |
| Asn | Cys | Ser | Gly 340 | Arg | Val | Glu | Val | Arg 345 | Ile | His | Glu | Gln | Trp 350 | Trp | Thr |
| Ile | Cys | Asp 355 | Gln | Asn | Trp | Lys | Asn 360 | Glu | Gln | Ala | Leu | Val 365 | Val | Cys | Lys |
| Gln | Leu 370 | Gly | Суз | Pro | Phe | Ser 375 | Val | Phe | Gly | Ser | Arg 380 | Arg | Ala | Lys | Pro |
| Ser 385 | Asn | Glu | Ala | Arg | Asp 390 | Ile | Trp | Ile | Asn | Ser 395 | Ile | Ser | Суз | Thr | Gly 400 |
| Asn | Glu | Ser | Ala | Leu 405 | Trp | Asp | Суѕ | Thr | Tyr 410 | Asp | Gly | Lys | Ala | Lys 415 | Arg |
| Thr | Cys | Phe | Arg 420 | Arg | Ser | Asp | Ala | Gly 425 | Val | Ile | Cys | Ser | Asp 430 | Lys | Ala |
| Ąsp | | Asp 435 | Leu | Arg | Leu | Val | Gly 440 | Ala | His | Ser | Pro | Cys 445 | Tyr | Gly | Arg |
| Leu | Glu 450 | Val | Lys | Tyr | Gln | Gly 455 | Glu | Trp | Gly | Thr | Val 460 | Суѕ | His | Asp | Arg |

| Trp 465 | Ser | Thr | Arg | Asn | Ala 470 | Ala | Val | Val | Суѕ | Lys 475 | Gln | Leu | Gly | Суз | Gly 480 |
|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|
| Lys | Pro | Met | His | Val 485 | Phe | Gly | Met | Thr | Tyr 490 | Phe | Lys | Glu | Ala | Ser 495 | Gly |
| Pro | Ile | Trp | Leu 500 | Asp | Asp | Val | Ser | Cys 505 | Ile | Gly | Asn | Glu | Ser 510 | Asn | Ile |
| Trp | Asp | Cys 515 | Glu | His | Ser | Gly | Trp 520 | Gly | Lys | His | Asn | Cys 525 | Val | His | Arg |
| Glu | Asp 530 | Val | Ile | Val | Thr | Cys 535 | Ser | Gly | Asp | Ala | Thr 540 | Trp | Gly | Leu | Arg |
| Leu 545 | Val | Gly | Gly | Ser | Asn 550 | Arg | Суз | Ser | Gly | Arg 555 | Leu | Glu | Val | Tyr | Phe 560 |
| Gln | Gly | Arg | Trp | Gly 565 | Thr | Val | Cys | Asp | Asp 570 | Gly | Trp | Asn | Ser | Lys 575 | Ala |
| Ala | Ala | Val | Val 580 | Cys | Ser | Gln | Leu | Asp 585 | Cys | Pro | Ser | Ser | Ile 590 | Ile | Gly |
| Met | Gly | Leu 595 | Gly | Asn | Ala | Ser | Thr 600 | Gly | Tyr | Gly | Lys | 11e 605 | Trp | Leu | Asp |
| _ | 610 | | | | | 615 | | | | | 620 | | Cys | | |
| 625 | | _ | | | 630 | | | | | 635 | | | Val | _ | 640 |
| | | | | 645 | | | | | 650 | | | | Gly | 655 | |
| | | | 660 | | | | | 665 | | | | | Ala 670 | | |
| | | 675 | | | | | 680 | | | | | 685 | Val | | - |
| | 690 | | | | | 695 | | | | | 700 | | G1u | | |
| Phe 705 | Thr | Glu | Arg | Thr | Leu 710 | His | Ile | Leu | Met | Ser 715 | Asn | Ser | Gly | Суѕ | Thr 720 |

| Gly | Gly | Glu | Ala | Ser 725 | Leu | Trp | Asp | Суз | Ile 730 | Arg | Trp | Glu | Trp | Lys 735 | Gln |
|-------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|
| Thr | Ala | Cys | His 740 | Leu | Asn | Met | Glu | Ala 745 | Ser | Leu | Ile | Cys | Ser 750 | Ala | His |
| Arg | Gln | Pro 755 | Arg | Leu | Val | Gly | Ala 760 | Asp | Met | Pro | Суз | Ser 765 | Gly | Arg | Val |
| Gl u | Val 770 | Lys | His | Ala | Asp | Thr 775 | Trp | Arg | Ser | Val | Cys 780 | Asp | Ser | Asp | Phe |
| Ser 785 | Leu | His | Ala | Ala | Asn 790 | Val | Leu | Cys | Arg | Glu 795 | Leu | Asn | Cys | Gly | 800 |
| Ala | Ile | Ser | Leu | Ser 805 | Val | Gly | Asp | His | Phe 810 | Gly | Lys | Gly | Asn | Gly 815 | Leu |
| Thr | Trp | Ala | Glu 820 | Lys | Phe | Gln | Cys | Glu 825 | Gly | Ser | Glu | Thr | His 830 | Leu | Ala |
| Leu | Cys | Pro 835 | Ile | Val | Gln | His | Pro 840 | Glu | Asp | Thr | Суѕ | Ile 845 | His | Ser | Arg |
| Glu | Val 850 | Gly | Val | Val | Суз | Ser 855 | Arg | Tyr | Thr | Asp | Val 860 | Arg | Leu | Val | Asn |
| Gly 865 | Lys | Ser | Gln | Суз | Asp 870 | Gly | Gln | Val | Glu | Ile 875 | neA | Val | Leu | Gly | His 880 |
| Trp | Gly | Ser | Leu | Cys 885 | Asp | Thr | His | Trp | Asp 890 | Pro | Glu | Asp | Ala | Arg 895 | Val |
| Leu | Суз | Arg | Gln 900 | Leu | Ser | Суѕ | Gly | Thr 905 | Ala | Leu | Ser | Thr | Thr 910 | Gly | Gly |
| ГÀЗ | Tyr | Ile 915 | Gly | Glu | Arg | Ser | Val 920 | Arg | Val | Trp | Gly | His 925 | Arg | Phe | His |
| Суз | Leu 930 | Gly | neA | Glu | Ser | Leu 935 | Leu | Asp | Asn | Суз | Gln 940 | Met | Thr | Val | Leu |
| Gly 945 | Ala | Pro | Pro | Cys | Ile 950 | His | Gly | neA | Thr | Val 955 | Ser | Val | Ile | Суз | Thr 960 |
| Gly | Ser | Leu | Thr | Gln 965 | Pro | Leu | Phe | Pro | Cys 970 | Leu | Ala | Asn | Val | Ser 975 | Asp |

Pro Tyr Leu Ser Ala Val Pro Glu Gly Ser Ala Leu Ile Cys Leu Glu 980 985 990

- Asp Lys Arg Leu Arg Leu Val Asp Gly Asp Ser Arg Cys Ala Gly Arg 995 1000 1005
- Val Glu Ile Tyr His Asp Gly Phe Trp Gly Thr Ile Cys Asp Asp Gly 1010 1015 1020
- Trp Asp Leu Ser Asp Ala His Val Val Cys Gln Lys Leu Gly Cys Gly 1025 1030 1035 1040
- Val Ala Phe Asn Ala Thr Val Ser Ala His Phe Gly Glu Gly Ser Gly
 1045 1050 1055
- Pro Ile Trp Leu Asp Asp Leu Asn Cys Thr Gly Thr Glu Ser His Leu 1060 1065 1070
- Trp Gln Cys Pro Ser Arg Gly Trp Gly Gln His Asp Cys Arg His Lys 1075 1080 1085
- Glu Asp Ala Gly Val Ile Cys Ser Glu Phe Thr Ala Leu Arg Leu Tyr . 1090 1095 1100
- Ser Glu Thr Glu Thr Glu Ser Cys Ala Gly Arg Leu Glu Val Phe Tyr 1105 1110 1115 1120
- Asn Gly Thr Trp Gly Ser Val Gly Arg Arg Asn Ile Thr Thr Ala Ile 1125 . 1130 1135
- Ala Gly Ile Val Cys Arg Gln Leu Gly Cys Gly Glu Asn Gly Val Val 1140 1145 1150
- Ser Leu Ala Pro Leu Ser Lys Thr Gly Ser Gly Phe Met Trp Val Asp 1155 1160 1165
- Asp Ile Gln Cys Pro Lys Thr His Ile Ser Ile Trp Gln Cys Leu Ser 1170 1175 1180
- Ala Pro Trp Glu Arg Arg Ile Ser Ser Pro Ala Glu Glu Thr Trp Ile 1185 1190 1195 1200
- Thr Cys Glu Asp Arg Ile Arg Val Arg Gly Gly Asp Thr Glu Cys Ser 1205 1210 1215
- Gly Arg Val Glu Ile Trp His Ala Gly Ser Trp Gly Thr Val Cys Asp 1220 1225 1230

Asp Ser Trp Asp Leu Ala Glu Ala Glu Val Val Cys Gln Gln Leu Gly 1235 1240 1245

Cys Gly Ser Ala Leu Ala Ala Leu Arg Asp Ala Ser Phe Gly Gln Gly 1250 1255 1260

Thr Gly Thr Ile Trp Leu Asp Asp Met Arg Cys Lys Gly Asn Glu Ser 1265 1270 1275 1280

Phe Leu Trp Asp Cys His Ala Lys Pro Trp Gly Gln Ser Asp Cys Gly 1285 1290 1295

His Lys Glu Asp Ala Gly Val Arg Cys Ser Gly Gln Ser Leu Lys Ser 1300 1305 1310

Leu Asn Ala Ser Ser Gly His Leu Ala Leu Ile Leu Ser Ser Ile Phe 1315 1320 1325

Gly Leu Leu Leu Val Leu Phe Ile Leu Phe Leu Thr Trp Cys Arg 1330 1335 1340

Val Gln Lys Gln Lys His Leu Pro Leu Arg Val Ser Thr Arg Arg Arg 1345 1350 1355 1360

Gly Ser Leu Glu Glu Asn Leu Phe His Glu Met Glu Thr Cys Leu Lys 1365 1370 1375

Arg Glu Asp Pro His Gly Thr Arg Thr Ser Asp Asp Thr Pro Asn His 1380 1385 1390

Gly Cys Glu Asp Ala Ser Asp Thr Ser Leu Leu Gly Val Leu Pro Ala 1395 1400 1405

Ser Glu Ala Thr Lys 1410

<210> 14

<211> 1319

<212> PRT

<213> Homo sapiens

<400> 14

Phe Asn Gly Thr Asp Leu Glu Leu Arg Leu Val Asn Gly Asp Gly Pro

1 5 10 15

Cys Ser Gly Thr Val Glu Val Lys Phe Gln Gly Gln Trp Gly Thr Val

WO 00/77239

Cys Asp Asp Gly Trp Asn Thr Thr Ala Ser Thr Val Val Cys Lys Gln Leu Gly Cys Pro Phe Ser Phe Ala Met Phe Arg Phe Gly Gln Ala Val Thr Arg His Gly Lys Ile Trp Leu Asp Asp Val Ser Cys Tyr Gly Asn Glu Ser Ala Leu Trp Glu Cys Gln His Arg Glu Trp Gly Ser His Asn Cys Tyr His Gly Glu Asp Val Gly Val Asn Cys Tyr Gly Glu Ala Asn Leu Gly Leu Arg Leu Val Asp Gly Asn Asn Ser Cys Ser Gly Arg Val Glu Val Lys Phe Gln Glu Arg Trp Gly Thr Ile Cys Asp Asp Gly Trp Asn Leu Asn Thr Ala Ala Val Val Cys Arg Gln Leu Gly Cys Pro Ser Ser Phe Ile Ser Ser Gly Val Val Asn Ser Pro Ala Val Leu Arg Pro Ile Trp Leu Asp Asp Ile Leu Cys Gln Gly Asn Glu Leu Ala Leu Trp

Asn Cys Arg His Arg Gly Trp Gly Asn His Asp Cys Ser His Asn Glu

Asp Val Thr Leu Thr Cys Tyr Asp Ser Ser Asp Leu Glu Leu Arg Leu

Val Gly Gly Thr Asn Arg Cys Met Gly Arg Val Glu Leu Lys Ile Gln

Gly Arg Trp Gly Thr Val Cys His His Lys Trp Asn Asn Ala Ala Ala

Asp Val Val Cys Lys Gln Leu Gly Cys Gly Thr Ala Leu His Phe Ala

Gly Leu Pro His Leu Gln Ser Gly Ser Asp Val Val Trp Leu Asp Gly

WO 00/77239

| 275 | 280 | 285 |
|-----|-----|-----|
|-----|-----|-----|

- Val Ser Cys Ser Gly Asn Glu Ser Phe Leu Trp Asp Cys Arg His Ser
- Gly Thr Val Asn Phe Asp Cys Leu His Gln Asn Asp Val Ser Val Ile
- Cys Ser Asp Gly Ala Asp Leu Glu Leu Arg Leu Ala Asp Gly Ser Asn
- Asn Cys Ser Gly Arg Val Glu Val Arg Ile His Glu Gln Trp Trp Thr
- Ile Cys Asp Gln Asn Trp Lys Asn Glu Gln Ala Leu Val Val Cys Lys
- Gln Leu Gly Cys Pro Phe Ser Val Phe Gly Ser Arg Arg Ala Lys Pro
- Ser Asn Glu Ala Arg Asp Ile Trp Ile Asn Ser Ile Ser Cys Thr Gly
- Asn Glu Ser Ala Leu Trp Asp Cys Thr Tyr Asp Gly Lys Ala Lys Arg
- Thr Cys Phe Arg Arg Ser Asp Ala Gly Val Ile Cys Ser Asp Lys Ala
- Asp Leu Asp Leu Arg Leu Val Gly Ala His Ser Pro Cys Tyr Gly Arg
- Leu Glu Val Lys Tyr Gln Gly Glu Trp Gly Thr Val Cys His Asp Arg
- Trp Ser Thr Arg Asn Ala Ala Val Val Cys Lys Gln Leu Gly Cys Gly
- Lys Pro Met His Val Phe Gly Met Thr Tyr Phe Lys Glu Ala Ser Gly
- Pro Ile Trp Leu Asp Asp Val Ser Cys Ile Gly Asn Glu Ser Asn Ile
- Trp Asp Cys Glu His Ser Gly Trp Gly Lys His Asn Cys Val His Arg
- Glu Asp Val Ile Val Thr Cys Ser Gly Asp Ala Thr Trp Gly Leu Arg

Leu Val Gly Gly Ser Asn Arg Cys Ser Gly Arg Leu Glu Val Tyr Phe Gln Gly Arg Trp Gly Thr Val Cys Asp Asp Gly Trp Asn Ser Lys Ala Ala Ala Val Val Cys Ser Gln Leu Asp Cys Pro Ser Ser Ile Ile Gly Met Gly Leu Gly Asn Ala Ser Thr Gly Tyr Gly Lys Ile Trp Leu Asp Asp Val Ser Cys Asp Gly Asp Glu Ser Asp Leu Trp Ser Cys Arg Asn Ser Gly Trp Gly Asn Asn Asp Cys Ser His Ser Glu Asp Val Gly Val Ile Cys Ser Asp Ala Ser Asp Met Glu Leu Arg Leu Val Gly Gly Ser Ser Arg Cys Ala Gly Lys Val Glu Val Asn Val Gln Gly Ala Val Gly Ile Leu Cys Ala Asn Gly Trp Gly Met Asn Ile Ala Glu Val Val Cys Arg Gln Leu Glu Cys Gly Ser Ala Ile Arg Val Ser Arg Glu Pro His Phe Thr Glu Arg Thr Leu His Ile Leu Met Ser Asn Ser Gly Cys Thr Gly Glu Ala Ser Leu Trp Asp Cys Ile Arg Trp Glu Trp Lys Gln Thr Ala Cys His Leu Asn Met Glu Ala Ser Leu Ile Cys Ser Ala His Arg Gln Pro Arg Leu Val Gly Ala Asp Met Pro Cys Ser Gly Arg Val Glu Val Lys His Ala Asp Thr Trp Arg Ser Val Cys Asp Ser Asp Phe

Ser Leu His Ala Ala Asn Val Leu Cys Arg Glu Leu Asn Cys Gly Asp

Ala Ile Ser Leu Ser Val Gly Asp His Phe Gly Lys Gly Asn Gly Leu Thr Trp Ala Glu Lys Phe Gln Cys Glu Gly Ser Glu Thr His Leu Ala Leu Cys Pro Ile Val Gln His Pro Glu Asp Thr Cys Ile His Ser Arg Glu Val Gly Val Val Cys Ser Arg Tyr Thr Asp Val Arg Leu Val Asn Gly Lys Ser Gln Cys Asp Gly Gln Val Glu Ile Asn Val Leu Gly His Trp Gly Ser Leu Cys Asp Thr His Trp Asp Pro Glu Asp Ala Arg Val Leu Cys Arg Gln Leu Ser Cys Gly Thr Ala Leu Ser Thr Thr Gly Gly 900 -Lys Tyr Ile Gly Glu Arg Ser Val Arg Val Trp Gly His Arg Phe His Cys Leu Gly Asn Glu Ser Leu Leu Asp Asn Cys Gln Met Thr Val Leu Gly Ala Pro Pro Cys Ile His Gly Asn Thr Val Ser Val Ile Cys Thr Gly Ser Leu Thr Gln Pro Leu Phe Pro Cys Leu Ala Asn Val Ser Asp Pro Tyr Leu Ser Ala Val Pro Glu Gly Ser Ala Leu Ile Cys Leu Glu Asp Lys Arg Leu Arg Leu Val Asp Gly Asp Ser Arg Cys Ala Gly Arg Val Glu Ile Tyr His Asp Gly Phe Trp Gly Thr Ile Cys Asp Asp Gly Trp Asp Leu Ser Asp Ala His Val Val Cys Gln Lys Leu Gly Cys Gly Val Ala Phe Asn Ala Thr Val Ser Ala His Phe Gly Glu Gly Ser Gly

WO 00/77239

Pro Ile Trp Leu Asp Asp Leu Asn Cys Thr Gly Thr Glu Ser His Leu

- Trp Gln Cys Pro Ser Arg Gly Trp Gly Gln His Asp Cys Arg His Lys
- Glu Asp Ala Gly Val Ile Cys Ser Glu Phe Thr Ala Leu Arg Leu Tyr
- Ser Glu Thr Glu Thr Glu Ser Cys Ala Gly Arg Leu Glu Val Phe Tyr
- Asn Gly Thr Trp Gly Ser Val Gly Arg Arg Asn Ile Thr Thr Ala Ile
- Ala Gly Ile Val Cys Arg Gln Leu Gly Cys Gly Glu Asn Gly Val Val
- Ser Leu Ala Pro Leu Ser Lys Thr Gly Ser Gly Phe Met Trp Val Asp
- Asp Ile Gln Cys Pro Lys Thr His Ile Ser Ile Trp Gln Cys Leu Ser
- Ala Pro Trp Glu Arg Arg Ile Ser Ser Pro Ala Glu Glu Thr Trp Ile
- Thr Cys Glu Asp Arg Ile Arg Val Arg Gly Gly Asp Thr Glu Cys Ser
- Gly Arg Val Glu Ile Trp His Ala Gly Ser Trp Gly Thr Val Cys Asp
- Asp Ser Trp Asp Leu Ala Glu Ala Glu Val Val Cys Gln Gln Leu Gly
- Cys Gly Ser Ala Leu Ala Leu Arg Asp Ala Ser Phe Gly Gln Gly
- Thr Gly Thr Ile Trp Leu Asp Asp Met Arg Cys Lys Gly Asn Glu Ser
- Phe Leu Trp Asp Cys His Ala Lys Pro Trp Gly Gln Ser Asp Cys Gly
- His Lys Glu Asp Ala Gly Val Arg Cys Ser Gly Gln Ser Leu Lys Ser

1300 1305 1310

Leu Asn Ala Ser Ser Gly His 1315

<210> 15

<211> 24

<212> PRT

<213> Homo sapiens

<400> 15

Leu Ala Leu Ile Leu Ser Ser Ile Phe Gly Leu Leu Leu Leu Val Leu

1 5 10 15

Phe Ile Leu Phe Leu Thr Trp Cys

20

<210> 16

<211> 70

<212> PRT

<213> Homo sapiens

<400> 16

Arg Val Gln Lys Gln Lys His Leu Pro Leu Arg Val Ser Thr Arg Arg

1 5 10 15

Arg Gly Ser Leu Glu Glu Asn Leu Phe His Glu Met Glu Thr Cys Leu 20 25 30

Lys Arg Glu Asp Pro His Gly Thr Arg Thr Ser Asp Asp Thr Pro Asn 35 40 45

His Gly Cys Glu Asp Ala Ser Asp Thr Ser Leu Leu Gly Val Leu Pro 50 55 60

Ala Ser Glu Ala Thr Lys 65 70

<210> 17

<211> 3104

<212> DNA

<213> Homo sapiens

<400> 17

gtcgacccac gcgtccggtc tgtggctgag catggccctc ccagccctgg gcctggaccc 60

| | | : | | | | 120 |
|------------|--------------------|------------|------------|------------|------------|------|
| | ctgggccttt | | | | | |
| | ggcgggcagg | | | | | |
| | agcttcttcc | | | | | |
| | aatactctct | | | | | |
| | gtccccaggc | | | | | |
| | gcctttaaga | | | | | |
| | tacaatgtca | | | | | |
| | ${\tt attgaacttc}$ | | | | | |
| | ggccaaagcc | | | | | |
| tgggatgctc | tattctggta | ctatgaacaa | cttcctgggc | agtgagccca | tcctgatgcg | 660 |
| cacactggga | tcccagcctg | tcctcaagac | cgacaacttc | ctccgctggc | tgcatcatga | 720 |
| cgcctccttt | gtggcagcca | tcccttcgac | ccaggtcgtc | tacttcttct | tcgaggagac | 780、 |
| agccagcgag | tttgacttct | ttgagaggct | ccacacatcg | cgggtggcta | gagtctgcaa | 840 |
| gaatgacgtg | ggcggcgaaa | agctgctgca | gaagaagtgg | accaccttcc | tgaaggccca | 900 |
| gctgctctgc | acccagccgg | ggcagctgcc | cttcaacgtc | atccgccacg | cggtcctgct | 960 |
| ccccgccgat | tctcccacag | ctcccacat | ctacgcagtc | ttcacctccc | agtggcaggt | 1020 |
| tggcgggacc | aggagetetg | cggtttgtgc | cttctctc | ttggacattg | aacgtgtctt | 1080 |
| | tacaaagagt | | | | | |
| | ccccggccag | | | | | |
| | gaccatttcc | | | | | |
| | gagtatacac | | | | | |
| | atgtacctgg | | | | | |
| | gctcatctgg | | | | | |
| | ctggccccca | | | | | |
| | cgagccaact | | | | | |
| | tgtgcctggg | | | | | |
| | tggaagcagg | | | | | |
| | aggageette | | | | | |
| | tccatcctgg | | | | | |
| | ggcccagcag | | | | | |
| | gtgcaggatg | | | | | |
| | cctgtgatct | | | | | |
| | gcaggcatcc | | | | | |
| | ctggctgccc | | | | | |
| | gtgctttcag | | | | | |
| | ggcaaggttc | | | | | |
| | caacacctcc | | | | | |
| | aactgcctag | | | | | |
| | acctggccat | | | | | |
| | aagaccacct | | | | | |
| | cagggtgatg | | | | | |
| | tctctaacag | | | | | |
| | | | | | | |
| | agaggatcct | | | | | |
| | tcctaaaaaa | | | | | |
| | tcatatgcta | | | | | |
| | cactcccttc | | | | | |
| | gcaccgctga | | | | | |
| ttgcttcagt | tggggcagac | tctgatccct | tctgccctgg | cagaatggca | ggggtaatct | 2940 |

gagcettett cacteettta eeetagetga eeeetteaee teteeeete eettteett 3000 tgttttggga tteagaaaae tgettgteag agactgttta tttttatta aaaatataag 3060 gettaaaaaa aaaaaaaaaa aaaaaaaaaa aaaagggegg eege 3104

<210> 18 <211> 2283 <212> DNA <213> Homo sapiens

<400> 18

atggccetce cagccetggg cetggacece tggagectee tgggcetttt cetettecaa 60 ctgcttcagc tgctgctgcc gacgacgacc gcggggggag gcgggcaggg gcccatgccc 120 agggtcagat actatgcagg ggatgaacgt agggcactta gettetteca ecagaaggge 180 ctccaggatt ttgacactct gctcctgagt ggtgatggaa atactctcta cgtgggggct 240 cgaqaaqcca ttctggcctt ggatatccag gatccagggg tccccaggct aaagaacatg 300 ataccqtqqc caqccaqtqa cagaaaaaag agtgaatgtg cctttaagaa gaagagcaat 360 gagacacagt gtttcaactt catcogtgtc ctggtttctt acaatgtcac ccatctctac 420 acctgcggca ccttcgcctt cagccctgct tgtaccttca ttgaacttca agattcctac 480 ctgttgccca tctcggagga caaggtcatg gagggaaaag gccaaagccc ctttgacccc 540 gctcacaagc atacggctgt cttggtggat gggatgctct attctggtac tatgaacaac 600 ttcctqqqca qtqaqcccat cctqatqcqc acactqqqat cccaqcctqt cctcaagacc 660 qacaacttcc tccgctgqct gcatcatgac gcctcctttg tggcagccat cccttcgacc 720 caggtcgtct acttcttctt cgaggagaca gccagcgagt ttgacttctt tgaggaggctc 780 cacacatcgc qqqtqqctaq aqtctqcaaq aatqacqtqq gcqqcgaaaa gctqctqcaq 840 aagaagtgga ccaccttcct gaaggcccag ctgctctgca cccagccggg gcagctgccc 900 ttcaacgtca tccgccacgc ggtcctgctc cccgccgatt ctcccacagc tccccacatc 960 tacgcagtct tcacctccca gtggcaggtt ggcgggacca ggagctctgc ggtttgtgcc 1020 ttctctctct tggacattga acgtgtcttt aaggggaaat acaaagagtt gaacaaagaa 1080 actteacgct ggactactta taggggccct gagaccaacc cccggccagg cagttgctca 1140 gtgggccct cctctgataa ggccctgacc ttcatgaagg accatttcct gatggatgag 1200 caaqtqqtqq qqacqccct qctqqtqaaa tctqqcqtqq agtatacacq qcttqcaqtg 1260 gagacagece agggeettga tgggeacage catettgtea tgtacetggg aaccaceaca 1320 gggtcgctcc acaaggctgt ggtaagtggg gacagcagtg ctcatctggt ggaagagatt 1380 cagetyttee etgaceetga acetyttege aacetycage tygeeceeae ecagygtyca 1440 gtgtttgtag gcttctcagg aggtgtctgg agggtgcccc gagccaactg tagtgtctat 1500 qaqagetgtg tgqaetgtgt cettgeeegg gaceeecact gtgeetggga ceetgagtee 1560 cgaacctgtt gcctcctgtc tgcccccaac ctgaactcct ggaagcagga catggagcgg 1620 gggaacccag agtgggcatg tgccagtggc cccatgagca ggagccttcg gcctcagagc 1680 cgcccgcaaa tcattaaaga agtcctggct gtccccaact ccatcctgga gctcccctgc 1740 ccccacctgt cagccttggc ctcttattat tggagtcatg gcccagcagc agtcccagaa 1800 qcctcttcca ctqtctacaa tggctccctc ttqctqatag tgcaggatgg agttgggggt 1860 ctctaccagt gctgggcaac tgagaatggc ttttcatacc ctgtgatctc ctactgggtg 1920 gacagecagg accagacect ggeeetggat cetgaaetgg caggeatece cegggageat 1980 gtgaaggtcc cgttgaccag ggtcagtggt ggggccgccc tggctgccca gcagtcctac 2040 tggccccact ttgtcactgt cactgtcctc tttgccttag tgctttcagg agccctcatc 2100 atecteqtqq cetececatt qaqaqcaete egqqeteqqq qeaaggttea qqqetqtqaq 2160 accetgegee etggggagaa ggeecegtta ageagagage aacaceteea gteteceaag 2220

gaatgcagga cctctgccag tgatgtggac gctgacaaca actgcctagg cactgaggta 2280 gct 2283

<210> 19

<211> 761

<212> PRT

<213> Homo sapiens

<400> 19

Met Ala Leu Pro Ala Leu Gly Leu Asp Pro Trp Ser Leu Leu Gly Leu
1 5 10 15

Phe Leu Phe Gln Leu Leu Gln Leu Leu Leu Pro Thr Thr Ala Gly
20 25 30

Gly Gly Gln Gly Pro Met Pro Arg Val Arg Tyr Tyr Ala Gly Asp 35 40 45

Glu Arg Arg Ala Leu Ser Phe Phe His Gln Lys Gly Leu Gln Asp Phe
50 55 60

Asp Thr Leu Leu Ser Gly Asp Gly Asn Thr Leu Tyr Val Gly Ala 65 70 75 80

Arg Glu Ala Ile Leu Ala Leu Asp Ile Gln Asp Pro Gly Val Pro Arg 85 90 95

Leu Lys Asn Met Ile Pro Trp Pro Ala Ser Asp Arg Lys Ser Glu 100 105 110

Cys Ala Phe Lys Lys Ser Asn Glu Thr Gln Cys Phe Asn Phe Ile 115 120 125

Arg Val Leu Val Ser Tyr Asn Val Thr His Leu Tyr Thr Cys Gly Thr 130 135 140

Phe Ala Phe Ser Pro Ala Cys Thr Phe Ile Glu Leu Gln Asp Ser Tyr 145 150 155 160

Leu Leu Pro Ile Ser Glu Asp Lys Val Met Glu Gly Lys Gly Gln Ser 165 170 175

Pro Phe Asp Pro Ala His Lys His Thr Ala Val Leu Val Asp Gly Met 180 185 190

Leu Tyr Ser Gly Thr Met Asn Asn Phe Leu Gly Ser Glu Pro Ile Leu 195 200 205

Met Arg Thr Leu Gly Ser Gln Pro Val Leu Lys Thr Asp Asn Phe Leu Arg Trp Leu His His Asp Ala Ser Phe Val Ala Ala Ile Pro Ser Thr Gln Val Val Tyr Phe Phe Glu Glu Thr Ala Ser Glu Phe Asp Phe Phe Glu Arg Leu His Thr Ser Arg Val Ala Arg Val Cys Lys Asn Asp Val Gly Glu Lys Leu Leu Gln Lys Lys Trp Thr Thr Phe Leu Lys Ala Gln Leu Leu Cys Thr Gln Pro Gly Gln Leu Pro Phe Asn Val Ile Arg His Ala Val Leu Leu Pro Ala Asp Ser Pro Thr Ala Pro His Ile Tyr Ala Val Phe Thr Ser Gln Trp Gln Val Gly Gly Thr Arg Ser Ser Ala Val Cys Ala Phe Ser Leu Leu Asp Ile Glu Arg Val Phe Lys Gly Lys Tyr Lys Glu Leu Asn Lys Glu Thr Ser Arg Trp Thr Thr Tyr Arg Gly Pro Glu Thr Asn Pro Arg Pro Gly Ser Cys Ser Val Gly Pro Ser Ser Asp Lys Ala Leu Thr Phe Met Lys Asp His Phe Leu Met Asp Glu Gln Val Val Gly Thr Pro Leu Leu Val Lys Ser Gly Val Glu Tyr Thr Arg Leu Ala Val Glu Thr Ala Gln Gly Leu Asp Gly His Ser His Leu Val Met Tyr Leu Gly Thr Thr Thr Gly Ser Leu His Lys Ala Val Val Ser Gly Asp Ser Ser Ala His Leu Val Glu Glu Ile Gln Leu Phe Pro

| Asp 465 | Pro | Glu | Pro | Val | Arg 470 | Asn | Leu | Gln | Leu | Ala 475 | Pro | Thr | Gln | Gly | Ala 480 |
|------------|------------|------------|------------|------------|------------|-------------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|
| Val | Phe | Val | Gly | Phe 485 | Ser | Gly | Gly | Val | Trp 490 | Arg | Val | Pro | Arg | Ala 495 | Asn |
| Суз | Ser | Val | Tyr 500 | Glu | Ser | Суз | Val | Asp 505 | Суѕ | Val | Leu | Ala | Arg 510 | Asp | Pro |
| His | Суз | Ala 515 | Trp | Asp | Pro | Glu | Ser 520 | Arg | Thr | Cys | Суз | Leu 525 | Leu | Ser | Ala |
| Pro | Asn 530 | Leu | Asn | Ser | Trp | Lys 535 | Gln | Asp | Met | Glu | Arg 540 | Gly | Asn | Pro | Glu |
| Trp 545 | Ala | Суз | Ala | Ser | Gly 550 | Pro | Met | Ser | Arg | Ser 555 | Leu | Arg | Pro | Gln | Ser 560 |
| Arg | Pro | Gln | Ile | Ile 565 | Lys | Glu | Val | Leu | Ala 570 | Val | Pro | Asn | Ser | Ile 575 | Leu |
| Glu | Leu | Pro | Cys 580 | Pro | His | Leu | Ser | Ala 585 | Leu | Ala | Ser | Tyr | Tyr 590 | Trp | Ser |
| | | 595 | | | Val | | 600 | | | | | 605 | | | _ |
| Ser | Leu 610 | Leu | Leu | Ile | Val | Gln 615 | Asp | Gly | Val | Gly | Gly 620 | Leu | Tyr | Gln | Cys |
| Trp 625 | Ala | Thr | Glu | Asn | Gly 630 | Phe | Ser | Tyr | Pro | Val 635 | Ile | Ser | Tyr | Trp | Val 640 |
| | | | | 645 | Thr | | | | 650 | | | | | 655 | |
| | | | 660 | | Lys | | | 665 | | | | | 670 | | |
| Ala | Leu | Ala 675 | Ala | Gln | Gln | Ser | Tyr 680 | Trp | Pro | His | Phe | Val 685 | Thr | Val | Thr |
| Val | Leu 690 | Phe | Ala | Leu | Val | Leu 695 | Ser | GŢÀ | Ala | Leu | Ile 700 | Ile | Leu | Val | Ala |
| Ser 705 | Pro | Leu | Arg | Ala | Leu 710 | Arg | Ala | Arg | Gly | Lys 715 | Val | Gln | Gly | Суз | Glu 720 |

Thr Leu Arg Pro Gly Glu Lys Ala Pro Leu Ser Arg Glu Gln His Leu 730 Gln Ser Pro Lys Glu Cys Arg Thr Ser Ala Ser Asp Val Asp Ala Asp 745 Asn Asn Cys Leu Gly Thr Glu Val Ala 755 760 <210> 20 <211> 31 <212> PRT <213> Homo sapiens <400> 20 Met Ala Leu Pro Ala Leu Gly Leu Asp Pro Trp Ser Leu Leu Gly Leu 1 Phe Leu Phe Gln Leu Leu Gln Leu Leu Leu Pro Thr Thr Ala 20 25 <210> 21 <211> 730 <212> PRT <213> Homo sapiens <400> 21 Gly Gly Gly Gln Gly Pro Met Pro Arg Val Arg Tyr Tyr Ala Gly 5 Asp Glu Arg Arg Ala Leu Ser Phe Phe His Gln Lys Gly Leu Gln Asp 20 25 30 Phe Asp Thr Leu Leu Ser Gly Asp Gly Asn Thr Leu Tyr Val Gly 35 40 Ala Arg Glu Ala Ile Leu Ala Leu Asp Ile Gln Asp Pro Gly Val Pro 50 55 Arg Leu Lys Asn Met Ile Pro Trp Pro Ala Ser Asp Arg Lys Lys Ser 65 70 75

85

Glu Cys Ala Phe Lys Lys Ser Asn Glu Thr Gln Cys Phe Asn Phe

| 11e | Arg | vaı | 100 | Val | Ser | Tyr | Asn | 105 | Thr | HIS | Leu | Tyr | 110 | Cys | : GTA |
|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|-------------------|------------|------------|
| Thr | Phe | Ala 115 | Phe | Ser | Pro | Ala | Cys 120 | Thr | Phe | Ile | Glu | Leu 125 | Gln | Asp | Ser |
| Tyr | Leu 130 | Leu | Pro | Ile | Ser | Glu 135 | Asp | Lys | Val | Met | Glu 140 | Gly | Lys | Gly | Gln |
| Ser 145 | Pro | Phe | Asp | Pro | Ala 150 | His | ГÀЗ | His | Thr | Ala 155 | Val | Leu | Val | Asp | Gly 160 |
| Met | Leu | Tyr | Ser | Gly 165 | Thr | Met | Asn | Asn | Phe 170 | Leu | Gly | Ser | Glu | Pro 175 | Ile |
| | | _ | 180 | | | | | 185 | | | | | 190 | | Phe |
| | _ | 195 | | | | | 200 | | | | | 205 | | | Ser |
| | 210 | | Val | | | 215 | | | | | 220 | | | | |
| 225 | | | Arg | | 230 | | | | | 235 | | | | | 240 |
| | | | Gly | 245 | | | | | 250 | | | | | 255 | |
| | | | Leu 260 | | | | | 265 | | | | | 270 | | |
| | _ | 275 | Ala | | | | 280 | | • | | | 285 | | | |
| | 290 | | Val | | | 295 | | • | | | 300 | | | | |
| Ser 305 | Ala | Val | Cys | Ala | Phe 310 | Ser | Leu | Leu | Asp | Ile 315 | Glu | Arg | Val | Phe | Lys 320 |
| Gly | Lys | Tyr | Lys | Glu 325 | Leu | Asn | Lys | Glu | Thr 330 | Ser | Arg | Trp | Thr | Thr 335 | Tyr |
| Arg | Gly | Pro | Glu 340 | Thr | Asn | Pro | Arg | Pro 345 | Gly | Ser | Cys | Ser | Val 350 | Gly | Pro |

Ser Ser Asp Lys Ala Leu Thr Phe Met Lys Asp His Phe Leu Met Asp Glu Gln Val Val Gly Thr Pro Leu Leu Val Lys Ser Gly Val Glu Tyr Thr Arg Leu Ala Val Glu Thr Ala Gln Gly Leu Asp Gly His Ser His Leu Val Met Tyr Leu Gly Thr Thr Gly Ser Leu His Lys Ala Val Val Ser Gly Asp Ser Ser Ala His Leu Val Glu Glu Ile Gln Leu Phe Pro Asp Pro Glu Pro Val Arq Asn Leu Gln Leu Ala Pro Thr Gln Gly Ala Val Phe Val Gly Phe Ser Gly Gly Val Trp Arg Val Pro Arg Ala Asn Cys Ser Val Tyr Glu Ser Cys Val Asp Cys Val Leu Ala Arg Asp Pro His Cys Ala Trp Asp Pro Glu Ser Arg Thr Cys Cys Leu Leu Ser Ala Pro Asn Leu Asn Ser Trp Lys Gln Asp Met Glu Arg Gly Asn Pro Glu Trp Ala Cys Ala Ser Gly Pro Met Ser Arg Ser Leu Arg Pro Gln Ser Arg Pro Gln Ile Ile Lys Glu Val Leu Ala Val Pro Asn Ser Ile 530° Leu Glu Leu Pro Cys Pro His Leu Ser Ala Leu Ala Ser Tyr Tyr Trp Ser His Gly Pro Ala Ala Val Pro Glu Ala Ser Ser Thr Val Tyr Asn Gly Ser Leu Leu Leu Ile Val Gln Asp Gly Val Gly Gly Leu Tyr Gln Cys Trp Ala Thr Glu Asn Gly Phe Ser Tyr Pro Val Ile Ser Tyr Trp

Val Asp Ser Gln Asp Gln Thr Leu Ala Leu Asp Pro Glu Leu Ala Gly 615 620 Ile Pro Arg Glu His Val Lys Val Pro Leu Thr Arg Val Ser Gly Gly 630 635 Ala Ala Leu Ala Ala Gln Gln Ser Tyr Trp Pro His Phe Val Thr Val 650 Thr Val Leu Phe Ala Leu Val Leu Ser Gly Ala Leu Ile Ile Leu Val 665 Ala Ser Pro Leu Arg Ala Leu Arg Ala Arg Gly Lys Val Gln Gly Cys 675 680 Glu Thr Leu Arg Pro Gly Glu Lys Ala Pro Leu Ser Arg Glu Gln His 690 695 Leu Gln Ser Pro Lys Glu Cys Arg Thr Ser Ala Ser Asp Val Asp Ala 705 710 715 Asp Asn Asn Cys Leu Gly Thr Glu Val Ala 730 725 <210> 22 <211> 652 <212> PRT <213> Homo sapiens <400> 22 Gly Gly Gly Gln Gly Pro Met Pro Arg Val Arg Tyr Tyr Ala Gly 5 10 Asp Glu Arg Arg Ala Leu Ser Phe Phe His Gln Lys Gly Leu Gln Asp 20 25 Phe Asp Thr Leu Leu Ser Gly Asp Gly Asn Thr Leu Tyr Val Gly Ala Arg Glu Ala Ile Leu Ala Leu Asp Ile Gln Asp Pro Gly Val Pro 55 Arg Leu Lys Asn Met Ile Pro Trp Pro Ala Ser Asp Arg Lys Lys Ser 70 75 Glu Cys Ala Phe Lys Lys Ser Asn Glu Thr Gln Cys Phe Asn Phe

39

90

| Ile | Arg | Val | Leu 100 | Val | Ser | Tyr | Asn | Val 105 | Thr | His | Leu | Tyr | Thr 110 | Cys | Gly |
|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|-------------|------------|------------|------------|-------------|------------|
| Thr | Phe | Ala 115 | Phe | Ser | Pro | Ala | Cys 120 | Thr | Phe | Ile | Glu | Leu 125 | Gln | Asp | Ser |
| Tyr | Leu 130 | Leu | Pro | Ile | Ser | Glu 135 | Asp | Lys | Val | Met | Glu 140 | Gly | Lys | Gly | Gln |
| Ser 145 | Pro | Phe | Asp | Pro | Ala 150 | His | Lys | His | Thr | Ala 155 | Val | Leu | Val | Asp | Gly 160 |
| Met | Leu | Tyr | Ser | Gly 165 | Thr | Met | Asn | Asn | Phe 170 | Leu | Gly | Ser | Glu | Pro 175 | Ile |
| Leu | Met | Arg | Thr 180 | Leu | Gly | Ser | Gln | Pro 185 | Val | Leu | Lys | Thr | Asp 190 | Asn | Phe |
| Leu | Arg | Trp 195 | Leu | His | His | qeA | Ala 200 | Ser | Phe | Val | Ala | Ala 205 | Ile | Pro | Ser |
| Thr | Gln 210 | Val | Val | Tyr | Phe | Phe 215 | Phe | Glu | Glu | Thr | Ala 220 | Ser | Glu | Phe | Asp |
| Phe 225 | Phe | Glu | Arg | Leu | His 230 | Thr | Ser | Arg | Val | Ala 235 | Arg | Val | Cys | Lys | Asn 240 |
| Asp | Val | Gly | Gly | Glu 245 | Lys | Leu | Leu | Gln | Lys 250 | Lys | Trp | Thr | Thr | Phe 255. | Leu |
| Lys | Ala | Gln | Leu 260 | Leu | Суз | Thr | Gln | Pro 265 | Gly | Gln | Leu | Pro | Phe 270 | Asn | Val |
| Ile | Arg | His 275 | Ala | Val | Leu | Leu | Pro 280 | Ala | Asp | Ser | Pro | Thr 285 | Ala | Pro | His |
| Ile | Tyr 290 | Ala | Val | Phe | Thr | Ser 295 | Gln | Trp | Gln | V al | Gly 300 | Gly | Thr | Arg | Ser |
| Ser 305 | Ala | Val | Суз | Ala | Phe 310 | Ser | Leu | Leu | Asp | Ile 315 | Glu | Arg | Val | Phe | Lys 320 |
| Gly | Lys | Tyr | Lys | Glu 325 | Leu | Asn | Lys | Glu | Thr 330 | Ser | Arg | Trp | Thr | Thr 335 | Tyr |
| Arg | Gly | Pro | Glu 340 | Thr | Asn | Pro | Arg | Pro 345 | Gly | Ser | Суз | Ser | Val 350 | Gly | Pro |

Ser Ser Asp Lys Ala Leu Thr Phe Met Lys Asp His Phe Leu Met Asp 355 360 365

- Glu Gln Val Val Gly Thr Pro Leu Leu Val Lys Ser Gly Val Glu Tyr 370 375 380
- Thr Arg Leu Ala Val Glu Thr Ala Gln Gly Leu Asp Gly His Ser His 385 390 395 400
- Leu Val Met Tyr Leu Gly Thr Thr Thr Gly Ser Leu His Lys Ala Val 405 410 415
- Val Ser Gly Asp Ser Ser Ala His Leu Val Glu Glu Ile Gln Leu Phe 420 425 430
- Pro Asp Pro Glu Pro Val Arg Asn Leu Gln Leu Ala Pro Thr Gln Gly
 435 440 445
 - Ala Val Phe Val Gly Phe Ser Gly Gly Val Trp Arg Val Pro Arg Ala 450 455 460
 - Asn Cys Ser Val Tyr Glu Ser Cys Val Asp Cys Val Leu Ala Arg Asp 465 470 475 480
 - Pro His Cys Ala Trp Asp Pro Glu Ser Arg Thr Cys Cys Leu Leu Ser 485 490 495
 - Ala Pro Asn Leu Asn Ser Trp Lys Gln Asp Met Glu Arg Gly Asn Pro 500 505 510
 - Glu Trp Ala Cys Ala Ser Gly Pro Met Ser Arg Ser Leu Arg Pro Gln 515 520 525
 - Ser Arg Pro Gln Ile Ile Lys Glu Val Leu Ala Val Pro Asn Ser Ile 530 535 540
 - Leu Glu Leu Pro Cys Pro His Leu Ser Ala Leu Ala Ser Tyr Tyr Trp 545 550 555 560
 - Ser His Gly Pro Ala Ala Val Pro Glu Ala Ser Ser Thr Val Tyr Asn 565 570 575
 - Gly Ser Leu Leu Leu Ile Val Gln Asp Gly Val Gly Gly Leu Tyr Gln 580 585 590
 - Cys Trp Ala Thr Glu Asn Gly Phe Ser Tyr Pro Val Ile Ser Tyr Trp 595 600 605

Val Asp Ser Gln Asp Gln Thr Leu Ala Leu Asp Pro Glu Leu Ala Gly 615 Ile Pro Arg Glu His Val Lys Val Pro Leu Thr Arg Val Ser Gly Gly 635 640 630 Ala Ala Leu Ala Ala Gln Gln Ser Tyr Trp Pro His 650 645 <210> 23 <211> 21 <212> PRT <213> Homo sapiens <400> 23 Phe Val Thr Val Thr Val Leu Phe Ala Leu Val Leu Ser Gly Ala Leu 15 Ile Ile Leu Val Ala 20 <210> 24 <211> 57 <212> PRT <213> Homo sapiens Ser Pro Leu Arg Ala Leu Arg Ala Arg Gly Lys Val Gln Gly Cys Glu - 5 10 15 Thr Leu Arg Pro Gly Glu Lys Ala Pro Leu Ser Arg Glu Gln His Leu 20 25 Gln Ser Pro Lys Glu Cys Arg Thr Ser Ala Ser Asp Val Asp Ala Asp 35 40 45 Asn Asn Cys Leu Gly Thr Glu Val Ala 55 <210> 25 <211> 2964

42

<212> DNA

<213> Homo sapiens

<400> 25 gtegacceae gegteegegg aegegtgggg aeggeteeeg getgeagtet geeegeeege 60 cccgcgcggg ggccgagtcg cgaagcgcgc ctgcgacccg gcgtccgggc gcgctggaga 120 ggacgcgagg agccatgagg cgccagcctg cgaaggtggc ggcgctgctg ctcgggctgc 180 tettggagtg cacagaagee aaaaageatt getggtattt egaaggaete tatecaacet 240 attatatatg ccgctcctac gaggactgct gtggctccag gtgctgtgtg cgggccctct 300 ccatacagag gctgtggtac ttctggttcc ttctgatgat gggcgtgctt ttctgctgcg 360 gageeggett etteateegg aggegeatgt acceeegge getgategag gageeageet 420 tcaatgtgtc ctacaccagg cagcccccaa atcccggccc aggagcccag cagccggggc 480 cgccctatta cactgaccca ggaggaccgg ggatgaaccc tgtcgggaat tccatggcaa 540 tggctttcca ggtcccaccc aactcacccc aggggagtgt ggcctgcccg cccctccag 600 cctactgcaa cacgcctccg cccccgtacg aacaggtagt gaaggccaag tagtggggtg 660 cccacgtgca agaggagaga caggagaggg cctttccctg gcctttctgt cttcgttgat 720 gttcacttcc aggaacggtc tcgtgggctg ctaagggcag ttcctctgat atcctcacag 780 caagcacagc tetettteag gettteeatg gagtacaata tatgaactea caetttgtet 840 cctctgttgc ttctgtttct gacgcagtct gtgctctcac atggtagtgt ggtgacagtc 900 eccgaggget gacgteetta eggtggegtg accagateta eaggagagag actgagagga 960 agaaggcagt gctggaggtg caggtggcat gtagaggggc caggccgagc atcccaggca 1020 agcatectte tgccegggta ttaataggaa gececatgee gggeggetea geegatgaag 1080 cagcagecga etgagetgag eccagcaggt catetgetee ageetgteet etegteagee 1140 tteetettee agaagetgtt ggagagacat teaggagaga geaageeeet tgteatgttt 1200 ctgtctctgt tcatatccta aagatagact tctcctgcac cgccagggaa gggtagcacg 1260 tgcagetete accgcaggat ggggeetaga atcaggettg cettggagge etgacagtga 1320 tetgacatec actaagcaaa tttatttaaa tteatgggaa ateaetteet geeccaaact 1380 gagacattgc attttgtgag ctcttggtct gatttggaga aaggactgtt acccattttt 1440 ttggtgtgtt tatggaagtg catgtagage gteetgeeet ttgaaateag aetgggtgtg 1500 tgtetteeet ggacateaet geeteteeag ggeattetea ggeeeggggg teteetteee 1560 tcaggcagct ccagtggtgg gttctgaagg gtgctttcaa aacggggcac atctggctgg 1620 gaagtcacat ggactettee agggagagag accagetgag gegtetetet etgaggttgt 1680 gttgggtcta agcgggtgtg tgctgggctc caaggaggag gagcttgctg ggaaaagaca 1740 ggagaagtac tgactcaact gcactgacca tgttgtcata attagaataa agaagaagtg 1800 gtcggaaatg cacattcctg gataggaatc acagctcacc ccaggatctc acaggtagtc 1860 teetgagtag ttgaeggeta geggggaget agtteegeeg catagttata gtgttgatgt 1920 gtgaacgctg acctgtcctg tgtgctaaga gctatgcagc ttagctgagg cgcctagatt 1980 actagatgtg ctgtatcacg gggaatgagg tgggggtgct tattttttaa tgaactaatc 2040 agageetett gagaaattgt taeteattga aetggageat caagacatet catggaagtg 2100 gatacggagt gatttggtgt ccatgctttt cactctgagg acatttaatc ggagaacctc 2160 ctggggaatt ttgtgggaga cacttgggaa caaaacagac accctgggaa tgcagttgca 2220 ageacagatg etgecaceag tgtetetgae eaccetggtg tgaetgetga etgecagegt 2280 ggtacctccc atgctgcagg cctccatcta aatgagacaa caaagcacaa tgttcactgt 2340 ttacaaccaa gacaactgeg tgggtccaaa cacteetett eetecaggte atttgttttg 2400 catttttaat gtctttattt tttgtaatga aaaagcacac taagctgccc ctggaatcgg 2460 gtgcagctga ataggcaccc aaaagtccgt gactaaattt cgtttgtctt tttgatagca 2520 aattatgtta agagacagtg atggctaggg ctcaacaatt ttgtattccc atgtttgtgt 2580 gagacagagt ttgttttccc ttgaacttgg ttagaattgt gctactgtga acgctgatcc 2640 . tgcatatgga agtcccactt tggtgacatt tcctggccat tcttgtttcc attgtgtgga 2700 tggtgggttg tgcccacttc ctggagtgag acagetectg gtgtgtagaa ttcccggage 2760

gtccgtggtt cagagtaaac ttgaagcaga tctgtgcatg cttttcctct gcaacaattg 2820

<210> 26

<211> 516

<212> DNA

<213> Homo sapiens

<400> 26

atgaggege agectgegaa ggtggeggeg etgetgete ggetgetett ggagtgeaca 60 gaagceaaaa ageattgetg gtattegaa ggactetate caacetatta tatatgeege 120 teetacgagg actgetgtg etecaggtge tgtgtgeggg eceteteea acagaggetg 180 tggtaettet ggtteettet gatgatggge gtgetttet getgeggage eggettette 240 ateeggagge geatgtacee eeggeeggg ategaggage eageetteaa tgtgteetae 300 aceaggeage eeceaaatee eggeeeagga geecaggage egggeegee etattacaet 360 gacceaggag gacegggat gaaceetgte gggaatteea tggeaatgge ttteeaggte 420 eeaceeaaet eaceecaggg gagtgtggee tgeeegeee etecageeta etgeaacaeg 480 eeteegeeee egtaegaaca ggtagtgaag geeaag

<210> 27

<211> 172

<212> PRT

<213> Homo sapiens

<400> 27

Met Arg Arg Gln Pro Ala Lys Val Ala Ala Leu Leu Leu Gly Leu Leu 1 5 10 15

Leu Glu Cys Thr Glu Ala Lys Lys His Cys Trp Tyr Phe Glu Gly Leu 20 25 30

Tyr Pro Thr Tyr Tyr Ile Cys Arg Ser Tyr Glu Asp Cys Cys Gly Ser
35 40 45

Arg Cys Cys Val Arg Ala Leu Ser Ile Gln Arg Leu Trp Tyr Phe Trp 50 55 60

Phe Leu Leu Met Met Gly Val Leu Phe Cys Cys Gly Ala Gly Phe Phe 65 70 75 80

Ile Arg Arg Met Tyr Pro Pro Pro Leu Ile Glu Glu Pro Ala Phe
85 90 95

Asn Val Ser Tyr Thr Arg Gln Pro Pro Asn Pro Gly Pro Gly Ala Gln 100 105 110

Gln Pro Gly Pro Pro Tyr Tyr Thr Asp Pro Gly Gly Pro Gly Met Asn 115 120 Pro Val Gly Asn Ser Met Ala Met Ala Phe Gln Val Pro Pro Asn Ser 135 Pro Gln Gly Ser Val Ala Cys Pro Pro Pro Pro Ala Tyr Cys Asn Thr 145 155 160 Pro Pro Pro Tyr Glu Gln Val Val Lys Ala Lys 165 170 <210> 28 <211> 22 <212> PRT <213> Homo sapiens Met Arg Arg Gln Pro Ala Lys Val Ala Ala Leu Leu Leu Gly Leu Leu 1 5 10 15 Leu Glu Cys Thr Glu Ala 20 <210> 29 <211> 150 <212> PRT <213> Homo sapiens Lys Lys His Cys Trp Tyr Phe Glu Gly Leu Tyr Pro Thr Tyr Tyr Ile 1 5 10 15 Cys Arg Ser Tyr Glu Asp Cys Cys Gly Ser Arg Cys Cys Val Arg Ala 20 25 Leu Ser Ile Gln Arg Leu Trp Tyr Phe Trp Phe Leu Leu Met Met Gly 40 45 Val Leu Phe Cys Cys Gly Ala Gly Phe Phe Ile Arg Arg Met Tyr 50 55 60 Pro Pro Pro Leu Ile Glu Glu Pro Ala Phe Asn Val Ser Tyr Thr Arg 65 70

Gln Pro Pro Asn Pro Gly Pro Gly Ala Gln Gln Pro Gly Pro Pro Tyr 85 90 Tyr Thr Asp Pro Gly Gly Pro Gly Met Asn Pro Val Gly Asn Ser Met 105 Ala Met Ala Phe Gln Val Pro Pro Asn Ser Pro Gln Gly Ser Val Ala 120 125 Cys Pro Pro Pro Pro Ala Tyr Cys Asn Thr Pro Pro Pro Pro Tyr Glu 135 140 Gln Val Val Lys Ala Lys 145 150 <210> 30 <211> 38 <212> PRT <213> Homo sapiens <400> 30 Lys Lys His Cys Trp Tyr Phe Glu Gly Leu Tyr Pro Thr Tyr Tyr Ile 10 Cys Arg Ser Tyr Glu Asp Cys Cys Gly Ser Arg Cys Cys Val Arg Ala 25 Leu Ser Ile Gln Arg Leu 35 <210> 31 <211> 21 <212> PRT <213> Homo sapiens Trp Tyr Phe Trp Phe Leu Leu Met Met Gly Val Leu Phe Cys Cys Gly 1 10 Ala Gly Phe Phe Ile 20

<210> 32 <211> 91 <212> PRT

<213> Homo sapiens

<400> 32

Arg Arg Arg Met Tyr Pro Pro Pro Leu Ile Glu Glu Pro Ala Phe Asn
1 5 10 15

Val Ser Tyr Thr Arg Gln Pro Pro Asn Pro Gly Pro Gly Ala Gln Gln
20 25 30

Pro Gly Pro Pro Tyr Tyr Thr Asp Pro Gly Gly Pro Gly Met Asn Pro
35 40 45

Val Gly Asn Ser Met Ala Met Ala Phe Gln Val Pro Pro Asn Ser Pro 50 55 60

Gln Gly Ser Val Ala Cys Pro Pro Pro Pro Ala Tyr Cys Asn Thr Pro 65 70 75 80

Pro Pro Pro Tyr Glu Gln Val Val Lys Ala Lys 85 90

<210> 33

<211> 1980

<212> DNA

<213> Homo sapiens

<400> 33

gtegacecae gegteegeag etttggacae tteetetget tgaggacaee ttgactaace 60 tccaagggca actaaaggat caagaaaggc ccagcacagc agaagatcag ctggatctag 120 ctcctgcagg agatgtgtac aaagacaatc ccagtcctct ggggatgttt cctcctgtgg 180 aatctctatg tctcatcctc tcagaccatt taccctggaa tcaaggcaag gattactcag 240 agggcacttg actatggtgt tcaagctgga atgaagatga ttgagcaaat gctaaaagaa 300 aagaaactcc cagatttaag cggttctgag tctcttgaat ttctaaaagt tgattatgta 360 aactacaatt tttcaaatat aaaaatcagt gccttttcat ttccaaatac ctcattggct 420 tttgtgcctg gagtgggaat caaagcgcta accaaccatg gcactgccaa catcagcaca 480 gactgggggt tcgagtctcc actttttgtt ctgtataact cctttgctga gcccatggag 540 aaacccattt taaagaactt aaatgaaatg ctctgtccca ttattgcaag tgaagtcaaa 600 gcgctaaatg ccaacctcag cacactggag gttttaacca agattgacaa ctacactctg 660 ctggattact ccctaatcag ttctccagaa attactgaga actaccttga cctgaacttg 720 aagggtgtat tctacccact ggaaaacctc accgaccccc ccttctcacc agttcctttt 780 gtgctcccag aacgcagcaa ctccatgctc tacattggaa tcgccgagta tttctttaaa 840 tctgcgtcct ttgctcattt cacagctggg gttttcaatc tcactctctc caccgaagag 900 atttccaacc attttgttca aaactctcaa ggccttggca acgtgctctc ccggattgca 960 gagatetaca tettgteeca gecetteatg gtgaggatea tggecacaga geeteecata 1020 atcaatctac aaccaggcaa tttcaccctg gacatccctg cctccatcat gatgctcacc 1080 caacccaaga actccacagt tgaaaccatc gtttccatgg acttcgttgc tagtaccagt 1140 gttggcctgg ttattttggg acaaagactg gtctgctcct tgtctctgaa cagattccgc 1200

<210> 34 <211> 1365 <212> DNA <213> Homo sapiens

<400> 34

atqtqtacaa agacaatccc aqtcctctqq qqatqtttcc tcctqtggaa tctctatqtc 60 tcatcctctc agaccattta ccctggaatc aaggcaagga ttactcagag ggcacttgac 120 tatggtgttc aagctggaat gaagatgatt gagcaaatgc taaaagaaaa gaaactccca 180 gatttaagcg gttctgagtc tcttgaattt ctaaaagttg attatgtaaa ctacaatttt 240 tcaaatataa aaatcagtgc cttttcattt ccaaatacct cattggcttt tgtgcctgga 300 gtgggaatca aagcgctaac caaccatggc actgccaaca tcagcacaga ctgggggttc 360 gagtetecae tittigtiet gtataactee titgetgage ceatggagaa acceatitta 420 aagaacttaa atgaaatgct ctgtcccatt attgcaagtg aagtcaaagc gctaaatgcc 480 aacetcagca cactggaggt tttaaccaag attgacaact acactctgct ggattactcc 540 ctaatcagtt ctccagaaat tactgagaac taccttgacc tgaacttgaa gggtgtattc 600 tacceactgg aaaacctcac cgacccccc ttctcaccag ttccttttgt gctcccagaa 660 cgcaqcaact ccatgctcta cattggaatc gccgagtatt tetttaaatc tgcgtccttt 720 qctcatttca cagctggggt tttcaatctc actctctcca ccgaagagat ttccaaccat 780 tttgttcaaa actctcaagg cettggcaac gtgctctccc ggattgcaga gatctacatc 840 ttqtcccaqc ccttcatqqt qaqqatcatq qccacaqaqc ctcccataat caatctacaa 900 ccaggcaatt tcaccctgga catccctgcc tccatcatga tgctcaccca acccaagaac 960 tecacagttg aaaccategt ttecatggae ttegttgeta gtaceagtgt tggeetggtt 1020 attttgggac aaagactggt ctgctccttg tctctgaaca gattccgcct tgctttgcca 1080 gagtccaatc gcagcaacat tgaggtcttg aggtttgaaa atattctatc gtccattctt 1140 cactttggag tecteceact ggecaatgea aaattgeage aaggatttee tetgeecaat 1200 ccacacaaat tottattogt caattoagat attgaagtto ttgagggttt cottttgatt 1260 tccaccgacc tgaagtatga aacatcctca aagcagcagc caagtttcca cgtatgggaa 1320 1365 ggtctgaacc tgataagcag acagtggagg gggaagtcag cccct

<210> 35 <211> 455

<212> PRT

<213> Homo sapiens

<400> 35

Met Cys Thr Lys Thr Ile Pro Val Leu Trp Gly Cys Phe Leu Leu Trp

1 5 10 15

Asn Leu Tyr Val Ser Ser Ser Gln Thr Ile Tyr Pro Gly Ile Lys Ala
20 25 30

Arg Ile Thr Gln Arg Ala Leu Asp Tyr Gly Val Gln Ala Gly Met Lys
35 40 45

Met Ile Glu Gln Met Leu Lys Glu Lys Lys Leu Pro Asp Leu Ser Gly
50 55 60

Ser Glu Ser Leu Glu Phe Leu Lys Val Asp Tyr Val Asn Tyr Asn Phe 65 70 75 80

Ser Asn Ile Lys Ile Ser Ala Phe Ser Phe Pro Asn Thr Ser Leu Ala 85 90 95

Phe Val Pro Gly Val Gly Ile Lys Ala Leu Thr Asn His Gly Thr Ala 100 105 110

Asn Ile Ser Thr Asp Trp Gly Phe Glu Ser Pro Leu Phe Val Leu Tyr 115 120 125

Asn Ser Phe Ala Glu Pro Met Glu Lys Pro Ile Leu Lys Asn Leu Asn 130 135 140

Glu Met Leu Cys Pro Ile Ile Ala Ser Glu Val Lys Ala Leu Asn Ala 145 150 155 160

Asn Leu Ser Thr Leu Glu Val Leu Thr Lys Ile Asp Asn Tyr Thr Leu 165 170 175

Leu Asp Tyr Ser Leu Ile Ser Ser Pro Glu Ile Thr Glu Asn Tyr Leu 180 185 190

Asp Leu Asn Leu Lys Gly Val Phe Tyr Pro Leu Glu Asn Leu Thr Asp 195 200 205

Pro Pro Phe Ser Pro Val Pro Phe Val Leu Pro Glu Arg Ser Asn Ser 210 215 220

Met Leu Tyr Ile Gly Ile Ala Glu Tyr Phe Phe Lys Ser Ala Ser Phe 225 230 235 240

Ala His Phe Thr Ala Gly Val Phe Asn Leu Thr Leu Ser Thr Glu Glu Ile Ser Asn His Phe Val Gln Asn Ser Gln Gly Leu Gly Asn Val Leu Ser Arg Ile Ala Glu Ile Tyr Ile Leu Ser Gln Pro Phe Met Val Arg Ile Met Ala Thr Glu Pro Pro Ile Ile Asn Leu Gln Pro Gly Asn Phe Thr Leu Asp Ile Pro Ala Ser Ile Met Met Leu Thr Gln Pro Lys Asn Ser Thr Val Glu Thr Ile Val Ser Met Asp Phe Val Ala Ser Thr Ser Val Gly Leu Val Ile Leu Gly Gln Arg Leu Val Cys Ser Leu Ser Leu Asn Arg Phe Arg Leu Ala Leu Pro Glu Ser Asn Arg Ser Asn Ile Glu Val Leu Arg Phe Glu Asn Ile Leu Ser Ser Ile Leu His Phe Gly Val Leu Pro Leu Ala Asn Ala Lys Leu Gln Gln Gly Phe Pro Leu Pro Asn Pro His Lys Phe Leu Phe Val Asn Ser Asp Ile Glu Val Leu Glu Gly Phe Leu Leu Ile Ser Thr Asp Leu Lys Tyr Glu Thr Ser Ser Lys Gln Gln Pro Ser Phe His Val Trp Glu Gly Leu Asn Leu Ile Ser Arg Gln

Trp Arg Gly Lys Ser Ala Pro

<210> 36 <211> 23

<212> PRT

<213> Homo sapiens

<400> 36

Met Cys Thr Lys Thr Ile Pro Val Leu Trp Gly Cys Phe Leu Leu Trp

1 5 10 15

Asn Leu Tyr Val Ser Ser Ser

20

<210> 37

<211> 432

<212> PRT

<213> Homo sapiens

<400> 37

Gln Thr Ile Tyr Pro Gly Ile Lys Ala Arg Ile Thr Gln Arg Ala Leu
1 5 10 15

Asp Tyr Gly Val Gln Ala Gly Met Lys Met Ile Glu Gln Met Leu Lys
20 25 30

Glu Lys Lys Leu Pro Asp Leu Ser Gly Ser Glu Ser Leu Glu Phe Leu
35 40 45

Lys Val Asp Tyr Val Asn Tyr Asn Phe Ser Asn Ile Lys Ile Ser Ala
50 55 60

Phe Ser Phe Pro Asn Thr Ser Leu Ala Phe Val Pro Gly Val Gly Ile
65 70 75 80

Lys Ala Leu Thr Asn His Gly Thr Ala Asn Ile Ser Thr Asp Trp Gly
85 90 95

Phe Glu Ser Pro Leu Phe Val Leu Tyr Asn Ser Phe Ala Glu Pro Met 100 105 110

Glu Lys Pro Ile Leu Lys Asn Leu Asn Glu Met Leu Cys Pro Ile Ile 115 120 125

Ala Ser Glu Val Lys Ala Leu Asn Ala Asn Leu Ser Thr Leu Glu Val 130 135 140

Leu Thr Lys Ile Asp Asn Tyr Thr Leu Leu Asp Tyr Ser Leu Ile Ser 145 150 155 160

Ser Pro Glu Ile Thr Glu Asn Tyr Leu Asp Leu Asn Leu Lys Gly Val 165 170 175

| Phe | Tyr | Pro | Leu 180 | Glu | Asn | Leu | Thr | Asp 185 | Pro | Pro | Phe | Ser | Pro 190 | Val | Pro |
|------------|------------|------------|------------|------------|-------------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|
| Phe | Val | Leu 195 | Pro | Glu | Arg · | Ser | Asn 200 | Ser | Met | Leu | Tyr | 11e 205 | Gly | Ile | Ala |
| Glu | Tyr 210 | Phe | Phe | Lys | Ser | Ala 215 | Ser | Phe | Ala | His | Phe 220 | Thr | Ala | Gly | Val |
| Phe 225 | Asn | Leu | Thr | Leu | Ser 230 | Thr | Glu | Glu | Ile | Ser 235 | Asn | His | Phe | Val | Gln 240 |
| Asn | Ser | Gln | Gly | Leu 245 | Gly | Asn | Val | Leu | Ser 250 | Arg | Ile | Ala | Glu | Ile 255 | Tyr |
| Ile | Leu | Ser | G1n 260 | Pro | Phe | Met | Val | Arg 265 | Ile | Met | Ala | Thr | G1u 270 | Pro | Pro |
| Ile | Ile | Asn 275 | Leu | Gln | Pro | Gly | Asn 280 | Phe | Thr | Leu | Asp | Ile 285 | Pro | Ala | Ser |
| Ile | Met 290 | Met | Leu | Thr | Gln | Pro 295 | Lys | Asn | Ser | Thr | Val 300 | Glu | Thr | Ile | Val |
| Ser 305 | Met | Asp | Phe | Val | Ala 310 | Ser | Thr | Ser | Val | Gly 315 | Leu | Val | Ile | Leu | Gly 320 |
| Gln | Arg | Leu | Val | Cys 325 | Ser | Leu | Ser | Leu | Asn 330 | Arg | Phe | Arg | Leu | Ala 335 | Leu |
| Pro | Glu | Ser | Asn 340 | Arg | Ser | Asn | Ile | Glu 345 | Val | Leu | Arg | Phe | G1u 350 | Asn | Ile |
| Leu | Ser | Ser 355 | Ile | Leu | His | Phe | Gly 360 | Val | Leu | Pro | Leu | Ala 365 | Asn | Ala | Lys |
| Leu | Gln 370 | Gln | Gly | Phe | Pro | Leu 375 | Pro | Asn | Pro | His | Lys 380 | Phe | Leu | Phe | Val |
| Asn 385 | Ser | Asp | Ile | Glu | Val 390 | Leu | Glu | Gly | Phe | Leu 395 | Leu | Ile | Ser | Thr | Asp 400 |
| Leu | Lys | Tyr | Glu | Thr 405 | Ser | Ser | Lys | Gln | Gln 410 | Pro | Ser | Phe | His | Vaļ 415 | Trp |
| Glu | Gly | Leu | Asn 420 | Leu | Ile | Ser | Arg | Gln 425 | Trp | Arg | Gly | Lys | Ser 430 | Ala | Pro |

PCT/US00/14858 WO 00/77239

<210> 38 <211> 483 <212> PRT <213> Homo sapiens <400> 38

Met Ala Arg Gly Pro Cys Asn Ala Pro Arg Trp Val Ser Leu Met Val 1 5 10

Leu Val Ala Ile Gly Thr Ala Val Thr Ala Ala Val Asn Pro Gly Val

Val Val Arg Ile Ser Gln Lys Gly Leu Asp Tyr Ala Ser Gln Gln Gly 40

Thr Ala Ala Leu Gln Lys Glu Leu Lys Arg Ile Lys Ile Pro Asp Tyr 55

Ser Asp Ser Phe Lys Ile Lys His Leu Gly Lys Gly His Tyr Ser Phe

Tyr Ser Met Asp Ile Arg Glu Phe Gln Leu Pro Ser Ser Gln Ile Ser 85 90

Met Val Pro Asn Val Gly Leu Lys Phe Ser Ile Ser Asn Ala Asn Ile 100 105 110

Lys Ile Ser Gly Lys Trp Lys Ala Gln Lys Arg Phe Leu Lys Met Ser 115 120 125

Gly Asn Phe Asp Leu Ser Ile Glu Gly Met Ser Ile Ser Ala Asp Leu 130 135

Lys Leu Gly Ser Asn Pro Thr Ser Gly Lys Pro Thr Ile Thr Cys Ser 145 150 155

Ser Cys Ser Ser His Ile Asn Ser Val His Val His Ile Ser Lys Ser 165 170

Lys Val Gly Trp Leu Ile Gln Leu Phe His Lys Lys Ile Glu Ser Ala 185

Leu Arg Asn Lys Met Asn Ser Gln Val Cys Glu Lys Val Thr Asn Ser 200

| V | al | 210 | Ser | гÀ2 | Leu | Gln | Pro 215 | Tyr | Phe | GIn | Thr | Leu 220 | Pro | Val | Met | Thi |
|---|----------|------------|------------|------------|------------|------------|------------|------------|---------------------|------------|------------|------------|------------|------------|------------|------------|
| | ys 25 | Ile | Asp | Ser | Val | Ala 230 | Gly | Ile | Asn | Tyr | Gly 235 | Leu | Val | Ala | Pro | Pro 240 |
| A | la | Thr | Thr | Ala | Glu 245 | Thr | Leu | Asp | Val | Gln 250 | Met | Lys | Gly | Glu | Phe 255 | Ту |
| S | er | Glu | Asn | His 260 | His | Asn | Pro | Pro | Pro 265 | Phe | Ala | Pro | Pro | Val 270 | Met | Glı |
| P | he | Pro | Ala 275 | Ala | His | Asp | Arg | Met 280 | Val | Tyr | Leu | Gly | Leu 285 | Ser | Asp | Туз |
| P | he | Phe 290 | Asn | Thr | Ala | Gly | Leu 295 | Val | Tyr | Gln | Glu | Ala 300 | Gly | Val | Leu | Lys |
| | et 05 | Thr | Leu | Arg | Asp | Asp 310 | Met | Ile | Pro | Lys | Glu 315 | Ser | Lys | Phe | Arg | Let 320 |
| T | hr | Thr | Lys | Phe | Phe 325 | Gly | Thr | Phe | Leu | Pro 330 | Glu | Val | Ala | Lys | Lys 335 | Phe |
| P | ro | Asn | Met | Lys 340 | Ile | Gln | Ile | His | Val 3 4 5 | Ser | Ala | Ser | Thr | Pro 350 | Pro | His |
| L | eu | Ser | Val 355 | Gln | Pro | Thr | Gly | Leu 360 | Thr | Phe | Tyr | Pro | Ala 365 | Val | Asp | Va] |
| G | ln | Ala 370 | Phe | Ala | Val | Leu | Pro 375 | Asn | Ser | Ser | Leu | Ala 380 | Ser | Leu | Phe | Let |
| | le 85 | Gly | Met | His | Thr | Thr 390 | Gly | Ser | Met | Glu | Val 395 | Ser | Ala | Glu | Ser | Asr 400 |
| A | rg | Leu | Val | Gly | Glu 405 | Leu | ГÀЗ | Leu | Asp | Arg 410 | Leu | Leu | Leu | Glu | Leu 415 | Lys |
| H | is | Ser | Asn | Ile 420 | Gly | Pro | Phe | Pro | Val 425 | Glu | Leu | Leu | Gln | Asp 430 | Ile | Met |
| A | sn | Tyr | Ile 435 | Val | Pro | Ile | Leu | Val 440 | Leu | Pro | Arg | Val | Asn 445 | Glu | Lys | Leu |
| G | ln | Lys 450 | Gly | Phe | Pro | Leu | Pro 455 | Thr | Pro | Ala | Arg | Val 460 | Gln | Leu | Tyr | Asn |

Val Val Leu Gln Pro His Gln Asn Phe Leu Leu Phe Gly Ala Asp Val 465 470 475 480

Val Tyr Lys

<210> 39

<211> 481

<212> PRT

<213> Homo sapiens

<400> 39

Met Gly Ala Leu Ala Arg Ala Leu Pro Ser Ile Leu Leu Ala Leu Leu 1 5 10 15

Leu Thr Ser Thr Pro Glu Ala Leu Gly Ala Asn Pro Gly Leu Val Ala
20 25 30

Arg Ile Thr Asp Lys Gly Leu Gln Tyr Ala Ala Gln Glu Gly Leu Leu · 35 40 45

Ala Leu Gln Ser Glu Leu Leu Arg Ile Thr Leu Pro Asp Phe Thr Gly
50 55 60

Asp Leu Arg Ile Pro His Val Gly Arg Gly Arg Tyr Glu Phe His Ser 65 70 75 80

Leu Asn Ile His Glu Phe Gln Leu Pro Ser Ser Gln Ile Ser Met Val
85 90 95

Pro Asn Val Gly Leu Lys Phe Ser Ile Ser Asn Ala Asn Ile Lys Ile 100 105 110

Ser Gly Lys Trp Lys Ala Gln Lys Arg Phe Leu Lys Met Ser Gly Asn 115 120 125

Phe Asp Leu Ser Ile Glu Gly Met Ser Ile Ser Ala Asp Leu Lys Leu 130 135 140

Gly Ser Asn Pro Thr Ser Gly Lys Pro Thr Ile Thr Cys Ser Ser Cys 145 150 155 160

Ser Ser His Ile Asn Ser Val His Val His Ile Ser Lys Ser Lys Val 165 170 175

Gly Trp Leu Ile Gln Leu Phe His Lys Lys Ile Glu Ser Ala Leu Arg

| | | | 180 | | | | | 185 | | | | | 190 | | |
|------------|------------|------------|------------|------------|------------|------------|------------|------------|--------------------|------------|------------|------------|-------------------------|------------|------------|
| Asn | Lys | Met 195 | Asn | Ser | Gln | Val | Cys 200 | Glu | Lys | Val | Thr | Asn 205 | Ser | Val | Ser |
| Ser | Lys 210 | Leu | Gln | Pro | Tyr | Phe 215 | Gln | Thr | Leu | Pro | Val 220 | Met | Thr | Lys | Ile |
| Asp 225 | Ser | Val | Ala | Gly | Ile 230 | Asn | Tyr | Gly | Leu | Val 235 | Ala | Pro | Pro | Ala | Thr 240 |
| Thr | Ala | Glu | Thr · | Leu 245 | Asp | Val | Gln | Met | Ly s 250 | Gly | Glu | Phe | Tyr | Ser 255 | Glu |
| Asn | His | His | Asn 260 | Pro | Pro | Pro | Phe | Ala 265 | Pro | Pro | Val | Met | Glu 270 | Phe | Pro |
| Ala | | His 275 | Asp | Arg | Met | Val | Tyr 280 | Leu | Gly | Leu | Ser | Asp 285 | Tyr | Phe | Phe |
| Asn | Thr 290 | Ala | Gly | Leu | Val | Tyr 295 | Gln | Glu | Ala | Gly | Val 300 | Leu | Lys | Met | Thr |
| Leu 305 | Arg | Asp | Asp | Met | Ile 310 | Pro | Lys | Glu | Ser | Lуз 315 | Phe | Arg | Leu | Thr | Thr 320 |
| Lys | Phe | Phe | Gly | Thr 325 | Phe | Leu | Pro | Glu | Val 330 | Ala | Lys | Lys | Phe | Pro 335 | Asn |
| Met | Lys | Ile | Gln 340 | Ile | His | Val | Ser | Ala 345 | Ser | Thr | Pro | Pro | His 350 | Leu | Ser |
| Val | Gln | Pro 355 | Thr | Gly | Leu | Thr | Phe 360 | Tyr | Pro | Ala | Val | Asp 365 | Val | Gln | Ala |
| Leu | Ala 370 | Val | Leu | Pro | Asn | Ser 375 | Ser | Leu | Ala | Ser | Leu 380 | Phe | Leu | Ile | Gly |
| Met 385 | His | Thr | Thr | Gly | Ser 390 | Met | Glu | Val | Ser | Ala 395 | Glu | Ser | Asn | Arg | Leu 400 |
| Val | Gly | Glu | Leu | Lys 405 | Leu | Asp | Arg | Leu | Leu 410 | Leu | Glu | Leu | Lys | His 415 | Ser |
| Asn | Ile | Gly | Pro 420 | Phe | Pro | Val | Glu | Leu 425 | Leu | Gln | Asp | Ile | Met [.] 430 | Asn | Tyr |
| Ile | Val | Pro | Ile | Leu | Val | Leu | Pro | Arg | Val | Asn | Glu | Lys | Leu | Gln | Lys |

435 440 445

Gly Phe Pro Leu Pro Thr Pro Ala Arg Val Gln Leu Tyr Asn Val Val 450 455 460

Leu Gln Pro His Gln Asn Phe Leu Leu Phe Gly Ala Asp Val Val Tyr 465 470 475 480

Lys

<210> 40

<211> 383

<212> PRT

<213> Caenorhabditis elegans

<400> 40

Met Arg Ile Ala His Ala Ser Ser Arg Gly Asn Ile Ser Ile Phe Ser 1 5 10 15

Val Phe Leu Ile Pro Leu Ile Ala Tyr Ile Leu Ile Leu Pro Gly Val 20 25 30

Arg Arg Lys Arg Val Val Thr Thr Val Thr Tyr Val Leu Met Leu Ala 35 40 45

Val Gly Gly Ala Leu Ile Ala Ser Leu Ile Tyr Pro Cys Trp Ala Ser 50 55 60

Gly Ser Gln Met Ile Tyr Thr Gln Phe Arg Gly His Ser Asn Glu Arg
65 70 75 80

Ile Leu Ala Lys Ile Gly Val Glu Ile Gly Leu Gln Lys Val Asn Val 85 90 95

Thr Leu Lys Phe Glu Arg Leu Leu Ser Ser Asn Asp Val Leu Pro Gly
100 105 110

Ser Asp Met Thr Glu Leu Tyr Tyr Asn Glu Gly Phe Asp Ile Ser Gly 115 120 125

Ile Ser Ser Met Ala Glu Ala Leu His His Gly Leu Glu Asn Gly Leu 130 135 140

Pro Tyr Pro Met Leu Ser Val Leu Glu Tyr Phe Ser Leu Asn Gln Asp 145 150 155 160

Ser Phe Asp Trp Gly Arg His Tyr Arg Val Ala Gly His Tyr Thr His Ala Ala Ile Trp Phe Ala Phe Ala Cys Trp Cys Leu Ser Val Val Leu Met Leu Phe Leu Pro His Asn Ala Tyr Lys Ser Ile Leu Ala Thr Gly Ile Ser Cys Leu Ile Ala Cys Leu Val Tyr Leu Leu Ser Pro Cys Glu Leu Arg Ile Ala Phe Thr Gly Glu Asn Phe Glu Arg Val Asp Leu Thr Ala Thr Phe Ser Phe Cys Phe Tyr Leu Ile Phe Ala Ile Gly Ile Leu Cys Val Leu Cys Gly Leu Gly Leu Gly Ile Cys Glu His Trp Arg Ile Tyr Thr Leu Ser Thr Phe Leu Asp Ala Ser Leu Asp Glu His Val Gly Pro Lys Trp Lys Lys Leu Pro Thr Gly Gly Pro Ala Leu Gln Gly Val Gln Ile Gly Ala Tyr Gly Thr Asn Thr Thr Asn Ser Ser Arg Asp Lys Asn Asp Ile Ser Ser Asp Lys Thr Ala Gly Ser Ser Gly Phe Gln Ser Arg Thr Ser Thr Cys Gln Ser Ser Ala Ser Ser Ala Ser Leu Arg Ser Gln Ser Ser Ile Glu Thr Val His Asp Glu Ala Glu Leu Glu Arg Thr His Val His Phe Leu Gln Glu Pro Cys Ser Ser Ser Ser Thr

<210> 41

<211> 399

<212> PRT

<213> Homo sapiens

| <4 | n | | 41 |
|-----------|----|---|------|
| \4 | vu | _ | 44.1 |

Met Lys Met Arg Phe Leu Gly Leu Val Val Cys Leu Val Leu Trp Pro 1 5 10 15

Leu His Ser Glu Gly Ser Gly Gly Lys Leu Thr Ala Val Asp Pro Glu
20 25 30

Thr Asn Met Asn Val Ser Glu Ile Ile Ser Tyr Trp Gly Phe Pro Ser 35 40 45

Glu Glu Tyr Leu Val Glu Thr Glu Asp Gly Tyr Ile Leu Cys Leu Asn $50 \hspace{1cm} 55 \hspace{1cm} 60$

Arg Ile Pro His Gly Arg Lys Asn His Ser Asp Lys Gly Pro Lys Pro 65 70 75 80

Val Val Phe Leu Gln His Gly Leu Leu Ala Asp Ser Ser Asn Trp Val 85 90 95

Thr Asn Leu Ala Asn Ser Ser Leu Gly Phe Ile Leu Ala Asp Ala Gly
100 105 110

Phe Asp Val Trp Met Gly Asn Ser Arg Gly Asn Thr Trp Ser Arg Lys 115 120 125

His Lys Thr Leu Ser Val Ser Gln Asp Glu Phe Trp Ala Phe Ser Tyr 130 135 140

Asp Glu Met Ala Lys Tyr Asp Leu Pro Ala Ser Ile Asn Phe Ile Leu 145 150 155 160

Asn Lys Thr Gly Gln Glu Gln Val Tyr Tyr Val Gly His Ser Gln Gly 165 170 175

Thr Thr Ile Gly Phe Ile Ala Phe Ser Gln Ile Pro Glu Leu Ala Lys 180 185 190

Arg Ile Lys Met Phe Phe Ala Leu Gly Pro Val Ala Ser Val Ala Phe 195 200 205

Cys Thr Ser Pro Met Ala Lys Leu Gly Arg Leu Pro Asp His Leu Ile 210 215 220

Lys Asp Leu Phe Gly Asp Lys Glu Phe Leu Pro Gln Ser Ala Phe Leu 225 230 235 240

Lys Trp Leu Gly Thr His Val Cys Thr His Val Ile Leu Lys Glu Leu 245 250 255

Cys Gly Asn Leu Cys Phe Leu Leu Cys Gly Phe Asn Glu Arg Asn Leu 260 265 270

Asn Met Ser Arg Val Asp Val Tyr Thr Thr His Ser Pro Ala Gly Thr 275 280 285

Ser Val Gln Asn Met Leu His Trp Ser Gln Ala Val Lys Phe Gln Lys 290 295 300

Phe Gln Ala Phe Asp Trp Gly Ser Ser Ala Lys Asn Tyr Phe His Tyr 305 310 315 320

Asn Gln Ser Tyr Pro Pro Thr Tyr Asn Val Lys Asp Met Leu Val Pro 325 330 335

Thr Ala Val Trp Ser Gly Gly His Asp Trp Leu Ala Asp Val Tyr Asp 340 345 350

Val Asn Ile Leu Leu Thr Gln Ile Thr Asn Leu Val Phe His Glu Ser 355 360 365

Ile Pro Glu Trp Glu His Leu Asp Phe Ile Trp Gly Leu Asp Ala Pro 370 375 380

Trp Arg Leu Tyr Asn Lys Ile Ile Asn Leu Met Arg Lys Tyr Gln 385 390 395

<210> 42

<211> 19

<212> PRT

<213> Mus sp.

<400> 42

Met Ala Pro Pro Ala Ala Arg Leu Ala Leu Leu Ser Ala Ala Ala Leu

1 5 10 15

Thr Leu Ala

<210> 43

<211> 451

<212> PRT

<213> Mus sp.

<400> 43

| Ala 1 | Arg | Pro | Ala | Pro 5 | Gly | Pro | Arg | Ser | Gly 10 | Pro | Glu | Суз | Phe | Thr 15 | Ala |
|------------|------------|------------|-------------|------------|------------|------------------|------------|------------|------------|------------|------------|------------|------------|------------------|------------|
| Asn | Gly | Ala | Asp 20 | Tyr | Arg | Gly | Thr | Gln 25 | Ser | Trp | Thr | Ala | Leu 30 | Gln | Gly |
| Glу | Lys | Pro 35 | Су ѕ | Leu | Phe | Trp | Asn 40 | Glu | Thr | Phe | Gln | His 45 | Pro | Tyr | Asn |
| Thr | Leu 50 | Lys | Tyr | Pro | Asn | Gly 55 | Glu | Gly | Gly | Leu | Gly 60 | Glu | His | Asn | Tyr |
| Cys 65 | Arg | Asn | Pro | Asp | Gly 70 | Asp | Val | Ser | Pro | Trp 75 | Суз | Tyr | Val | Ala | Glu 80 |
| His | Glu | Asp | Gly | Val 85 | Tyr | Trp | Lys | Туг | 90 CAa | Glu | Ile | Pro | Ala | Cys 95 | Gln |
| Met | Pro | Gly | Asn 100 | Leu | Gly | Суз | Tyr | Lys 105 | Asp | His | Gly | Asn | Pro 110 | Pro | Pro |
| Leu | Thr | Gly 115 | Thr | Ser | Lys | Thr | Ser 120 | Asn | Lys | Leu | Thr | Ile 125 | Gln | Thr | Суѕ |
| Ile | Ser 130 | Phe | Cys | Arg | Ser | Gln 135 | Arg | Phe | Lys | Phe | Ala 140 | Gly | Met | Glu | Ser |
| Gly 145 | Tyr | Ala | Суз | Phe | Cys 150 | Gly | Asn | Asn | Pro | Asp 155 | Tyr | Trp | Lys | His | Gly 160 |
| Glu | Ala | Ala | Ser | Thr 165 | Glu | Cys _. | Asn | Ser | Val 170 | Суз | Phe | Gly | Asp | His 175 | Thr |
| Gln | Pro | Суз | Gly 180 | Gly | Asp | Gly | Arg | Ile 185 | Ile | Leu | Phe | Asp | Thr 190 | Leu | Val |
| Gly | Ala | Cys 195 | Gly | Gly | Asn | Tyr | Ser 200 | Ala | Met | Ala | Ala | Val 205 | Val | Tyr | Ser |
| Pro | Asp 210 | Phe | Pro | Asp | Thr | Tyr 215 | Ala | Thr | Gly | Arg | Val 220 | Суз | Tyr | Trp | Thr |
| 11e 225 | Arg | Val | Pro | Gly | Ala 230 | Ser | Arg | Ile | His | Phe 235 | Asn | Phe | Thr | Leu | Phe 240 |
| Asp | Ile | Arg | Asp | Ser 245 | Ala | Asp | Met | Val | Glu 250 | Leu | Leu | Asp | Gly | Tyr 255 | Thr |

His Arg Val Leu Val Arg Leu Ser Gly Arg Ser Arg Pro Pro Leu Ser 260 265 Phe Asn Val Ser Leu Asp Phe Val Ile Leu Tyr Phe Phe Ser Asp Arg 280 Ile Asn Gln Ala Gln Gly Phe Ala Val Leu Tyr Gln Ala Thr Lys Glu 300 295 Glu Pro Pro Gln Glu Arg Pro Ala Val Asn Gln Thr Leu Ala Glu Val 310 315 Ile Thr Glu Gln Ala Asn Leu Ser Val Ser Ala Ala His Ser Ser Lys 325 330 Val Leu Tyr Val Ile Thr Pro Ser Pro Ser His Pro Pro Gln Thr Ala 340 345 Gln Val Ala Ile Pro Gly His Arg Gln Leu Gly Pro Thr Ala Thr Glu 355 360 Trp Lys Asp Gly Leu Cys Thr Ala Trp Arg Pro Ser Ser Ser Gln 375 380 Ser Gln Gln Leu Ser Gln Arg Phe Phe Cys Met Ser His Leu Asn Leu 390 395 Ile Glu Ser Leu His Gln Glu Thr Leu Gly Thr Val Val Ser Leu Gly 405 410 Leu Leu Glu Ile Ser Gly Pro Phe Ser Met Asn Leu Pro Leu Gln Ser 425 Pro Ser Leu Arg Arg Ser Ser Arg Val Arg Val Asn Lys Met Thr Ala 440 Ile Pro Ser 450 <210> 44 <211> 150 <212> PRT <213> Mus sp. <400> 44

62

5

Lys Lys His Cys Trp Tyr Phe Glu Gly Leu Tyr Pro Thr Tyr Tyr Ile

Cys Arg Ser Tyr Glu Asp Cys Cys Gly Ser Arg Cys Cys Val Arg Ala 25 Leu Ser Ile Gln Arg Leu Trp Tyr Phe Trp Phe Leu Leu Met Met Gly 40 Val Leu Phe Cys Cys Gly Ala Gly Phe Phe Ile Arg Arg Met Tyr 50 Pro Pro Pro Leu Ile Glu Glu Pro Thr Phe Asn Val Ser Tyr Thr Arg 65 70 75 80 Gln Pro Pro Asn Pro Ala Pro Gly Ala Gln Gln Met Gly Pro Pro Tyr 85 90 Tyr Thr Asp Pro Gly Gly Pro Gly Met Asn Pro Val Gly Asn Thr Met 100 . 105 Ala Met Ala Phe Gln Val Gln Pro Asn Ser Pro His Gly Gly Thr Thr 115 120 125 Tyr Pro Pro Pro Pro Ser Tyr Cys Asn Thr Pro Pro Pro Pro Tyr Glu 135 140 Gln Val Val Lys Asp Lys <210> 45 <211> 2044 <212> DNA <213> Homo sapiens <400> 45 gtcgacccac gcgtccgggg aattgcagca ggaaaatatg tgaagagttt ttaaacccac 60 aaattettet taetttagaa ttagttgtta cattggcagg aaaaaataaa tgcagatgtt 120 ggaccatgtt ggaaaccttg tcaagacagt ggattgtctc acacagaatg gaaatgtggc 180 ttctgattct ggtggcgtat atgttccaga gaaatgtgaa ttcagtacat atgccaacta 240 aagetgtgga eecagaagea tteatgaata ttagtgaaat cateeaacat caaggetate 300 cctgtgagga atatgaagtc gcaactgaag atgggtatat cctttctgtt aacaggattc 360 ctcgaggcct agtgcaacct aagaagacag gttccaggcc tgtggtgtta ctgcagcatg 420 geetagttgg aggtgetage aactggattt eeaacetgee eaacaatage etgggettea 480 ttctggcaga tgctggtttt gacgtgtgga tqqqqaacag caggggaaac qcctqqtctc 540 gaaaacacaa gacactetee atagaceaag atgagttetg ggettteagt tatgatgaga 600 tggctaggtt tgaccttcct gcagtgataa actttatttt gcagaaaacg ggccaggaaa 660 agatetatta tgteggetat teaeagggea eeaceatggg etttattgea tttteeacea 720

tgccagagct ggctcagaaa atcaaaatgt attttgcttt agcacccata gccactgtta 780

```
agcatgcaaa aagccccggg accaaatttt tgttgctgcc agatatgatg atcaagggat 840
tqtttggcaa aaaagaattt ctqtatcaga ccagatttct cagacaactt gttatttacc 900
tttgtggcca ggtgattctt gatcagattt gtagtaatat catgttactt ctgggtggat 960
tcaacaccaa caatatgaac atgagccgag caagtgtata tgctgcccac actcttgctg 1020
gaacatctgt gcaaaatatt ctacactgga gccaggcagt gaattctggt gaactccggg 1080
catttgactg ggggagtgag accaaaaatc tggaaaaatg caatcagcca actcctgtaa 1140
ggtacagagt cagagatatg acggtcccta cagcaatgtg gacaggaggt caggactggc 1200
tttcaaatcc agaagacgtg aaaatgctgc tctctgaggt gaccaacctc atctaccata 1260
agaatattcc tgaatgggct cacgtggatt tcatctgggg tttggatgct cctcaccgta 1320
tgtacaatga aatcatccat ctgatgcagc aggaggagac caacctttcc cagggacggt 1380
qtqaqqccgt attgtgaagc atctgacact gacgatctta ggacaacctc ctgagggatg 1440
gggctaggac ccatgaaggc agaattacgg agagcagaga cctagtatac atttttcaga 1500
ttccctgcac ttggcactaa atccgacact tacatttaca tttttttct gtaaattaaa 1560
gtacttatta ggtaaataga ggttttgtat gctattatat attctaccat cttgaagggt 1620
aggttttacc tgatagccag aaaatatcta gacattctct atatcattca ggtaaatctc 1680
tttaaaacac ctattgtttt ttctataagc catatttttg gagcactaaa gtaaaatggc 1740
aaattgggac agatattgag gtctggagtc tgtggattat tgttgacttt gacaaaataa 1800
gctagacatt ttcaccttgt tgccacagag acataacact acctcaggaa gctgagctgc 1860
tttaaggaca acaacaacaa aatcagtgtt acagtatgga tgaaatctat gttaagcatt 1920
ctcagaataa ggccaagttt tatagttgca tctcagggaa gaaaatttta taggatgttt 1980
ccqc
                                                               2044
```

<210> 46 <211> 1269 <212> DNA

<213> Homo sapiens

<400> 46

atgttggaaa ccttgtcaag acagtggatt gtctcacaca gaatggaaat gtggcttctg 60 attetggtgg cgtatatgtt ccagagaaat gtgaattcag tacatatgcc aactaaagct 120 gtggacccag aagcattcat gaatattagt gaaatcatcc aacatcaagg ctatcctgt 180 gaggaatatg aagtcgcaac tgaagatggg tatatccttt ctgttaacag gattcctcga 240 ggcctagtgc aacctaagaa gacaggttcc aggcctgtgg tgttactgca gcatggccta 300 gttggaggtg ctagcaactg gatttccaac ctgcccaaca atagcctggg cttcattctg 360 gcagatgctg gttttgacgt gtggatgggg aacagcaggg gaaacgcctg gtctcgaaaa 420 cacaagacac tetecataga ecaagatgag ttetgggett teagttatga tgaqatgget 480 aggtttgacc ttcctgcagt gataaacttt attttgcaga aaacgggcca ggaaaagatc 540 tattatgtcg getattcaca gggcaccacc atgggettta ttgcattttc caccatgcca 600 gagetggete agaaaateaa aatgtatttt getttageac eeatageeac tqttaaqeat 660 gcaaaaagcc ccgggaccaa atttttgttg ctgccagata tgatgatcaa gggattgttt 720 ggcaaaaaag aatttetgta teagaceaga ttteteagae aacttgttat ttacetttgt 780 ggccaggtga ttcttgatca gatttgtagt aatatcatgt tacttctggg tggattcaac 840 accaacaata tgaacatgag ccgagcaagt gtatatgctg cccacactct tgctggaaca 900 tetgtgcaaa atattetaca etggagecag geagtgaatt etggtgaact eegggeattt 960 gactggggga gtgagaccaa aaatctggaa aaatgcaatc agccaactcc tqtaaqqtac 1020 agagtcagag atatgacggt ccctacagca atgtggacag gaggtcagga ctggctttca 1080

aatccagaag acgtgaaaat gctgctctct gaggtgacca acctcatcta ccataagaat 1140 attcctgaat gggctcacgt ggatttcatc tggggtttgg atgctcctca ccgtatgtac 1200 aatgaaatca tccatctgat gcagcaggag gagaccaacc tttcccaggg acggtgtgag 1260 gccgtattg 1269

<210> 47

<211> 423

<212> PRT

<213> Homo sapiens

<400> 47

Met Leu Glu Thr Leu Ser Arg Gln Trp Ile Val Ser His Arg Met Glu

1 5 10 15

Met Trp Leu Leu Ile Leu Val Ala Tyr Met Phe Gln Arg Asn Val Asn 20 25 30

Ser Val His Met Pro Thr Lys Ala Val Asp Pro Glu Ala Phe Met Asn 35 40 45

Ile Ser Glu Ile Ile Gln His Gln Gly Tyr Pro Cys Glu Glu Tyr Glu
50 55 60

Val Ala Thr Glu Asp Gly Tyr Ile Leu Ser Val Asn Arg Ile Pro Arg
65 70 75 80

Gly Leu Val Gln Pro Lys Lys Thr Gly Ser Arg Pro Val Val Leu Leu 85 90 95

Gln His Gly Leu Val Gly Gly Ala Ser Asn Trp Ile Ser Asn Leu Pro 100 105 110

Asn Asn Ser Leu Gly Phe Ile Leu Ala Asp Ala Gly Phe Asp Val Trp
115 120 125

Met Gly Asn Ser Arg Gly Asn Ala Trp Ser Arg Lys His Lys Thr Leu 130 135 140

Ser Ile Asp Gln Asp Glu Phe Trp Ala Phe Ser Tyr Asp Glu Met Ala 145 150 155 160

Arg Phe Asp Leu Pro Ala Val Ile Asn Phe Ile Leu Gln Lys Thr Gly 165 170 175

Gln Glu Lys Ile Tyr Tyr Val Gly Tyr Ser Gln Gly Thr Thr Met Gly 180 185 190

Phe Ile Ala Phe Ser Thr Met Pro Glu Leu Ala Gln Lys Ile Lys Met Tyr Phe Ala Leu Ala Pro Ile Ala Thr Val Lys His Ala Lys Ser Pro Gly Thr Lys Phe Leu Leu Pro Asp Met Met Ile Lys Gly Leu Phe Gly Lys Lys Glu Phe Leu Tyr Gln Thr Arg Phe Leu Arg Gln Leu Val Ile Tyr Leu Cys Gly Gln Val Ile Leu Asp Gln Ile Cys Ser Asn Ile Met Leu Leu Gly Gly Phe Asn Thr Asn Asn Met Asn Met Ser Arg Ala Ser Val Tyr Ala Ala His Thr Leu Ala Gly Thr Ser Val Gln Asn Ile Leu His Trp Ser Gln Ala Val Asn Ser Gly Glu Leu Arg Ala Phe Asp Trp Gly Ser Glu Thr Lys Asn Leu Glu Lys Cys Asn Gln Pro Thr Pro Val Arg Tyr Arg Val Arg Asp Met Thr Val Pro Thr Ala Met Trp Thr Gly Gly Gln Asp Trp Leu Ser Asn Pro Glu Asp Val Lys Met Leu Leu Ser Glu Val Thr Asn Leu Ile Tyr His Lys Asn Ile Pro Glu Trp Ala His Val Asp Phe Ile Trp Gly Leu Asp Ala Pro His Arg Met Tyr Asn Glu Ile Ile His Leu Met Gln Gln Glu Glu Thr Asn Leu Ser Gln

Gly Arg Cys Glu Ala Val Leu

<210> 48 <211> 33

<212> PRT

<213> Homo sapiens

<400> 48

Met Leu Glu Thr Leu Ser Arg Gln Trp Ile Val Ser His Arg Met Glu

1 5 10 15

Met Trp Leu Leu Ile Leu Val Ala Tyr Met Phe Gln Arg Asn Val Asn 20 25 30

Ser

<210> 49

<211> 390

<212> PRT

<213> Homo sapiens

<400> 49

Val His Met Pro Thr Lys Ala Val Asp Pro Glu Ala Phe Met Asn Ile
1 5 10 15

Ser Glu Ile Ile Gln His Gln Gly Tyr Pro Cys Glu Glu Tyr Glu Val 20 25 30

Ala Thr Glu Asp Gly Tyr Ile Leu Ser Val Asn Arg Ile Pro Arg Gly
35 40 45

Leu Val Gln Pro Lys Lys Thr Gly Ser Arg Pro Val Val Leu Leu Gln
50 55 60

His Gly Leu Val Gly Gly Ala Ser Asn Trp Ile Ser Asn Leu Pro Asn 65 70 75 80

Asn Ser Leu Gly Phe Ile Leu Ala Asp Ala Gly Phe Asp Val Trp Met 85 90 95

Gly Asn Ser Arg Gly Asn Ala Trp Ser Arg Lys His Lys Thr Leu Ser 100 105 110

Ile Asp Gln Asp Glu Phe Trp Ala Phe Ser Tyr Asp Glu Met Ala Arg 115 120 125

Phe Asp Leu Pro Ala Val Ile Asn Phe Ile Leu Gln Lys Thr Gly Gln 130 135 140

Glu Lys Ile Tyr Tyr Val Gly Tyr Ser Gln Gly Thr Thr Met Gly Phe

| 145 | | | | | 150 | | | | | 155 | | | | | 160 |
|--------------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|
| Ile | Ala | Phe | Ser | Thr 165 | Met | Pro | Glu | Leu | Ala 170 | Gln | Lys | Ile | Lys | Met 175 | Tyr |
| Phe | Ala | Leu | Ala 180 | Pro | Ile | Ala | Thr | Val 185 | Lys | His | Ala | Lys | Ser 190 | Pro | Gly |
| Thr | Lys | Phe 195 | Leu | Leu | Leu | Pro | Asp 200 | Met | Met | Ile | Lys | Gly 205 | Leu | Phe | Gly |
| Lys | Lys 210 | Glu | Phe | Leu | Tyr | Gln 215 | Thr | Arg | Phe | Leu | Arg 220 | Gln | Leu | Val | Ile |
| Tyr 225 | Leu | Cys | Gly | Gln | Val 230 | Ile | Leu | Asp | Gln | Ile 235 | Суз | Ser | Asn | Ile | Met 240 |
| Leu | Leu | Leu | Gly | Gly 245 | Phe | Asn | Thr | Asn | Asn 250 | Met | Asn | Met | Ser | Arg 255 | Ala |
| Ser | Val | Tyr | Ala 260 | Ala | His | Thr | Leu | Ala 265 | Gly | Thr | Ser | Val | Gln 270 | Asn | Ile |
| Leu | His | Trp 275 | Ser | Gln | Ala | Val | Asn 280 | Ser | Gly | Glu | Leu | Arg 285 | Ala | Phe | Asp |
| Trp | Gly 290 | Ser | Glu | Thr | Lys | Asn 295 | Leu | Glu | Lys | Суѕ | Asn 300 | Gln | Pro | Thr | Pro |
| V al 305 | Arg | Tyr | Arg | Val | Arg 310 | Asp | Met | Thr | Val | Pro 315 | Thr | Ala | Met | Trp | Thr 320 |
| Gly | Gly | Gln | Asp | Trp 325 | Leu | Ser | Asn | Pro | Glu 330 | Asp | Val | Lys | Met | Leu 335 | Leu |
| Ser | Glu | Val | Thr 340 | Asn | Leu | Ile | Tyr | His 345 | Lys | Asn | Ile | Pro | Glu 350 | Trp | Ala |
| His | Val | Asp 355 | Phe | Ile | Trp | Gly | Leu 360 | Asp | Ala | Pro | His | Arg 365 | Met | Tyr | Asn |
| Glu | Ile 370 | Ile | His | Leu | Met | Gln 375 | Gln | Glu | Glu | Thr | Asn 380 | Leu | Ser | Gln | Gly |
| Arg 385 | Суз | Glu | Ala | Val | Leu 390 | | | | | | | | | | |

<210> 50 <211> 221 <212> PRT <213> Homo sapiens <400> 50 Val His Met Pro Th

Val His Met Pro Thr Lys Ala Val Asp Pro Glu Ala Phe Met Asn Ile

Ser Glu Ile Ile Gln His Gln Gly Tyr Pro Cys Glu Glu Tyr Glu Val 20 25 30

Ala Thr Glu Asp Gly Tyr Ile Leu Ser Val Asn Arg Ile Pro Arg Gly
35 40 45

Leu Val Gln Pro Lys Lys Thr Gly Ser Arg Pro Val Val Leu Leu Gln 50 55 60

His Gly Leu Val Gly Gly Ala Ser Asn Trp Ile Ser Asn Leu Pro Asn 65 70 75 80

Asn Ser Leu Gly Phe Ile Leu Ala Asp Ala Gly Phe Asp Val Trp Met 85 90 95

Gly Asn Ser Arg Gly Asn Ala Trp Ser Arg Lys His Lys Thr Leu Ser 100 105 110

Ile Asp Gln Asp Glu Phe Trp Ala Phe Ser Tyr Asp Glu Met Ala Arg 115 120 125

Phe Asp Leu Pro Ala Val Ile Asn Phe Ile Leu Gln Lys Thr Gly Gln 130 135 140

Glu Lys Ile Tyr Tyr Val Gly Tyr Ser Gln Gly Thr Thr Met Gly Phe 145 150 155 160

Ile Ala Phe Ser Thr Met Pro Glu Leu Ala Gln Lys Ile Lys Met Tyr 165 170 175

Phe Ala Leu Ala Pro Ile Ala Thr Val Lys His Ala Lys Ser Pro Gly 180 185 190

Thr Lys Phe Leu Leu Pro Asp Met Met Ile Lys Gly Leu Phe Gly
195 200 205

Lys Lys Glu Phe Leu Tyr Gln Thr Arg Phe Leu Arg Gln 210 215 220

10

<210> 51
<211> 25
<212> PRT
<213> Homo sapiens
<400> 51
Leu Val Ile Tyr Leu Cys Gly Gln Val Ile Leu Asp Gln Ile Cys Ser

Asn Ile Met Leu Leu Gly Gly Phe 20 25

<210> 52 <211> 144

<212> PRT

<213> Homo sapiens

<400> 52

Asn Thr Asn Asn Met Asn Met Ser Arg Ala Ser Val Tyr Ala Ala His 1 5 10 15

Thr Leu Ala Gly Thr Ser. Val Gln Asn Ile Leu His Trp Ser Gln Ala 20 25 30

Val Asn Ser Gly Glu Leu Arg Ala Phe Asp Trp Gly Ser Glu Thr Lys 35 40 45

Asn Leu Glu Lys Cys Asn Gln Pro Thr Pro Val Arg Tyr Arg Val Arg 50 55 60

Asp Met Thr Val Pro Thr Ala Met Trp Thr Gly Gly Gln Asp Trp Leu 65 70 75 80

Ser Asn Pro Glu Asp Val Lys Met Leu Leu Ser Glu Val Thr Asn Leu 85 90 95

Ile Tyr His Lys Asn Ile Pro Glu Trp Ala His Val Asp Phe Ile Trp
100 105 110

Gly Leu Asp Ala Pro His Arg Met Tyr Asn Glu Ile Ile His Leu Met 115 120 125

Gln Gln Glu Glu Thr Asn Leu Ser Gln Gly Arg Cys Glu Ala Val Leu 130 135 140

<210> 53 <211> 2133 <212> DNA <213> Homo sapiens

<400> 53

qtcqacccac qcqtccacqq cqaqqqctcc cqqqqqcqcaq cattqccccc cctqcaccac 60 ctcaccaaga tggctacttt gggacacaca ttccccttct atgctggccc caagccaacc 120 ttcccgatgg acaccacttt ggccagcatc atcatgatct ttctgactgc actggccacg 180 ttcatcgtca tcctgcctgg cattcgggga aagacgaggc tgttctggct gcttcgggtg 240 gtgaccaget tatteategg ggetgeaate etggetgtga attteagtte tgagtggtet 300 gtgggccagg tcagcaccaa cacatcatac aaggccttca gttctgagtg gatcagcgct 360 gatattgggc tgcaggtcgg gctgggtgga gtcaacatca cactcacagg gacccccgtg 420 cagcagetga atgagaceat caattacaae gaggagttea eetggegeet gggtgagaae 480 tatgctgagg agtgtgcaaa ggctctggag aaggggctgc cagaccctgt gttgtaccta 540 gctgagaagt tcactccaag aagcccatgt ggcctatacc gccagtaccg cctggcggga 600 cactacacct cagccatgct atgggtggca ttcctctgct ggctgctggc caatgtgatg 660 ctctccatgc ctgtgctggt atatggtggc tacatgctat tggccacggg catcttccag 720 etgttggete tgetettett etceatggee acateactea ceteaceetg teceetgeae 780 etgggegett etgtgetgea tacteaceat gggeetgeet tetggateae attgaceaea 840 ggactgctgt gtgtgctgct gggcctggct atggcggtgg cccacaggat gcagcctcac 900 aggetgaagg etttetteaa ecagagtgtg gatgaagace ecatgetgga gtggagteet 960 gaggaaggtg gacteetgag ecceegetae eggteeatgg etgaeagtee caagteecag 1020 gacatteece tqteaqagge tteeteeace aaggeatact gtaaggagge acaceecaaa 1080 gatectgatt gtgetttata acattectee eegtggagge eacetggaet teeagtetgg 1140 ctccaaacct cattggcgcc ccataaaacc agcagaactg ccctcagggt ggctgttacc 1200 agacacccag caccaatcta cagacggagt agaaaaagga ggctctatat actgatgtta 1260 aaaaacaaaa caaaacaaaa agceetaagg gactgaagag atgetgggee tgteeataaa 1320 geotyttgee atgataagge caageagggg etagettate tgeacageaa eecageettt 1380 ccgtgctgcc ttgcctcttc aagatgctat tcactgaaac ctaacttcac ccccataaca 1440 ccagcagggt gggggttaca tatgattctc ctatggtttc ctctcatccc tcggcacctc 1500 ttgttttcct ttttcctggg ttccttttgt tcttccttta cttctccage ttgtgtggcc 1560 ttttggtaca atgaaagaca gcactggaaa ggaggggaaa ccaaacttct catcctaggt 1620 ctaacattaa ccaactatgc cacattetet ttgagettea gtteecaaat ttgetacata 1680 agattgcaag acttgccaag aatcttggga tttatctttc tatgccttgc tgacacctac 1740 cttggccctc aaacaccacc tcacaagaag ccaggtggga agttagggaa tcaactccaa 1800 aacgctattc cttcccaccc cactcagctg ggctagctga gtggcatcca ggacggggga 1860 gtgggtgacc tgcctcatca ctgccaccta acgtccccct ggggtggttc agaaagatgc 1920 tagetetggt agggteeete eggeeteaet agagggegee eetattaete tggagtegae 1980 gcagagaatc aggtttcaca gcactgcgga gagtgtacta ggctgtctcc agcccagcga 2040 ageteatgag gacgtgegae eeeggegegg agaaqeeatg aaaattaatg ggaaaaacag 2100 tttttaaaaa aaaaaaaaaa aaagggcggc cgc 2133

<211> 1029

<212> DNA <213> Homo sapiens <400> 54 atggctactt tgggacacac atteceette tatgetggee ceaagecaac ettecegatg 60 qacaccactt tggccagcat catcatgate tttctgactg cactggccac gttcatcgte 120 atcctgcctg gcattcgggg aaagacgagg ctgttctggc tgcttcgggt ggtgaccagc 180 ttattcatcg gggctgcaat cctggctgtg aatttcagtt ctgagtggtc tgtgggccag 240 qtcaqcacca acacatcata caaqqccttc aqttctqaqt qqatcaqcqc tqatattqqq 300 ctgcaggtcg ggctgggtgg agtcaacatc acactcacag ggacccccgt gcagcagctg 360 aatgagacca tcaattacaa cgaggagttc acctggcgcc tgggtgagaa ctatgctgag 420 gagtgtgcaa aggctctgga gaaggggctg ccagaccctg tgttgtacct agctgagaag 480 ttcactccaa gaageccatg tggcctatac cgccagtace gcctggcggg acactacacc 540 tcagccatgc tatgggtggc attcctctgc tggctgctgg ccaatgtgat gctctccatg 600 cctgtgctgg tatatggtgg ctacatgcta ttggccacgg gcatcttcca gctgttggct 660 etgetettet tetecatgge cacateacte aceteaceet gteccetgea cetgggeget 720 tetgtgetge atacteacca tgggeetgee ttetggatea cattgaccae aggaetgetg 780 tgtgtgctgc tgggcctggc tatggcggtg gcccacagga tgcagcctca caggctgaag 840 getttettea accagagtgt ggatgaagae eecatgetgg agtggagtee tgaggaaggt 900 qqacteetqa qeeeceqeta eeqqteeatq qetqacaqte eeaaqteeca qqacatteee 960 ctgtcagagg cttcctccac caaggcatac tgtaaggagg cacaccccaa agatcctgat 1020 tgtgcttta <210> 55 <211> 343 <212> PRT <213> Homo sapiens <400> 55 Met Ala Thr Leu Gly His Thr Phe Pro Phe Tyr Ala Gly Pro Lys Pro 5 15 1 10 Thr Phe Pro Met Asp Thr Thr Leu Ala Ser Ile Ile Met Ile Phe Leu 20 25 30 Thr Ala Leu Ala Thr Phe Ile Val Ile Leu Pro Gly Ile Arg Gly Lys 40 Thr Arg Leu Phe Trp Leu Leu Arg Val Val Thr Ser Leu Phe Ile Gly 55 Ala Ile Leu Ala Val Asn Phe Ser Ser Glu Trp Ser Val Gly Gln 70 Val Ser Thr Asn Thr Ser Tyr Lys Ala Phe Ser Ser Glu Trp Ile Ser 85 90

1029

| Ala | Asp | Ile | Gly 100 | Leu | Gln | Val | Gly | Leu 105 | Gly | Gly | Val | Asn | Ile 110 | Thr | Leu |
|------------|------------|------------|------------|------------|------------|------------|-------------------|------------|------------|------------|------------|------------|------------|------------|------------|
| Thr | Gly | Thr 115 | Pro | Val | Gln | Gln | Leu 120 | Asn | Glu | Thr | Ile | Asn 125 | Tyr | Asn | Glu |
| Glu | Phe 130 | Thr | Trp | Arg | Leu | Gly 135 | Glu | Asn | Туг | Ala | Glu 140 | Glu | Суѕ | Ala | Lys |
| Ala 145 | Leu | Glu | Lys | Gly | Leu 150 | Pro | Asp | Pro | Val | Leu 155 | туг | Leu | Ala | Glu | Lys 160 |
| Phe | Thr | Pro | Arg | Ser 165 | Pro | Суз | Gly | Leu | Tyr 170 | Arg | Gln | Tyr | Arg | Leu 175 | Ala |
| Gly | His | Tyr | Thr 180 | Ser | Ala | Met | Leu | Trp 185 | Val | Ala | Phe | Leu | Cys 190 | Trp | Leu |
| Leu | Ala | Asn 195 | Val | Met | Leu | Ser | Met 200 | Pro | Val | Leu | Val | Tyr 205 | Gly | Gly | Tyr |
| Met | Leu 210 | Leu | Ala | Thr | Gly | Ile 215 | Phe | Gln | Leu | Leu | Ala 220 | Leu | Leu | Phe | Phe |
| Ser 225 | Met | Ala | Thr | Ser | Leu 230 | Thr | Ser | Pro | Суѕ | Pro 235 | Leu | His | Leu | Gly | Ala 240 |
| Ser | Val | Leu | His | Thr 245 | His | His | Gly | Pro | Ala 250 | Phe | Trp | Ile | Thr | Leu 255 | Thr |
| Thr | Gly | Leu | Leu 260 | Суз | Val | Leu | Leu | Gly 265 | Leu | Ala | Met | Ala | Val 270 | Ala | His |
| Arg | Met | Gln 275 | Pro | His | Arg | Leu | Lys 280 | Ala | Phe | Phe | Asn | Gln 285 | Ser | Val | Asp |
| Glu | Asp 290 | Pro | Met | Leu | Glu | Trp 295 | Ser | Pro | Glu | Glu | 300 Gly | Gly | Leu | Leu | Ser |
| Pro 305 | Arg | Tyr | Arg | Ser | Met 310 | Ala | Asp | Ser | Pro | 315 Tys | Ser | Gln | Asp | Ile | Pro 320 |
| Leu | Ser | Glu | Ala | Ser 325 | Ser | Thr | Lys | Ala | Tyr 330 | Суз | Lys | Glu | Ala | His 335 | Pro |

73

Lys Asp Pro Asp Cys Ala Leu 340

<210> 56 <211> 23 <212> PRT <213> Homo sapiens <400> 56 Met Ala Thr Leu Gly His Thr Phe Pro Phe Tyr Ala Gly Pro Lys Pro 10 5 Thr Phe Pro Met Asp Thr Thr 20 <210> 57 <211> 112 <212> PRT <213> Homo sapiens <400> 57 Asn Phe Ser Ser Glu Trp Ser Val Gly Gln Val Ser Thr Asn Thr Ser 5 10 Tyr Lys Ala Phe Ser Ser Glu Trp Ile Ser Ala Asp Ile Gly Leu Gln 20 25 Val Gly Leu Gly Gly Val Asn Ile Thr Leu Thr Gly Thr Pro Val Gln 40 Gln Leu Asn Glu Thr Ile Asn Tyr Asn Glu Glu Phe Thr Trp Arg Leu - 55 Gly Glu Asn Tyr Ala Glu Glu Cys Ala Lys Ala Leu Glu Lys Gly Leu 70 **7**5 Pro Asp Pro Val Leu Tyr Leu Ala Glu Lys Phe Thr Pro Arg Ser Pro 85 90 95

<210> 58

100

<211> 22

Cys Gly Leu Tyr Arg Gln Tyr Arg Leu Ala Gly His Tyr Thr Ser Ala

105

<212> PRT <213> Homo sapiens <400> 58 Thr Ser Leu Thr Ser Pro Cys Pro Leu His Leu Gly Ala Ser Val Leu His Thr His His Gly Pro 20 <210> 59 <211> 19 <212> PRT <213> Homo sapiens <400> 59 Leu Ala Ser Ile Ile Met Ile Phe Leu Thr Ala Leu Ala Thr Phe Ile - 10 Val Ile Leu <210> 60 <211> 20 <212> PRT <213> Homo sapiens <400> 60 Leu Phe Trp Leu Leu Arg Val Val Thr Ser Leu Phe Ile Gly Ala Ala 10 Ile Leu Ala Val <210> 61 <211> 22 <212> PRT <213> Homo sapiens <400> 61 Met Leu Trp Val Ala Phe Leu Cys Trp Leu Leu Ala Asn Val Met Leu 10

75

Ser Met Pro Val Leu Val 20

```
<210> 62
<211> 17
<212> PRT
<213> Homo sapiens
<400> 62
Leu Ala Thr Gly Ile Phe Gln Leu Leu Ala Leu Leu Phe Phe Ser Met
                                    10
Ala
<210> 63
<211> 22
<212> PRT
<213> Homo sapiens
<400> 63
Ala Phe Trp Ile Thr Leu Thr Thr Gly Leu Leu Cys Val Leu Leu Gly
                  5
                                     10
                                                         15
Leu Ala Met Ala Val Ala
             20
<210> 64
<211> 8
<212> PRT
<213> Homo sapiens
<400> 64
Pro Gly Ile Arg Gly Lys Thr Arg
<210> 65
<211> 6
<212> PRT
<213> Homo sapiens
<400> 65
Tyr Gly Gly Tyr Met Leu
 1
                5
```

76 .

<210> 66 <211> 72

<212> PRT <213> Homo sapiens <400> 66 His Arg Met Gln Pro His Arg Leu Lys Ala Phe Phe Asn Gln Ser Val 10 5 Asp Glu Asp Pro Met Leu Glu Trp Ser Pro Glu Glu Gly Leu Leu 25 20 Ser Pro Arg Tyr Arg Ser Met Ala Asp Ser Pro Lys Ser Gln Asp Ile Pro Leu Ser Glu Ala Ser Ser Thr Lys Ala Tyr Cys Lys Glu Ala His 50 55 60 Pro Lys Asp Pro Asp Cys Ala Leu 70 65 <210> 67 <211> 4928 <212> DNA <213> Mus sp. <400> 67 qteqaeceae qegteegeee ggeteeeggt getgeeeeet etgeeeeggg eegegeeegg 60 gggtcccgca ctgacggccc atggcgccgc ccgccgcccg tctcgcgctg ctctccgccg 120 ctgcgctcac tctggcggcc cggcccgcgc ccggtccccg ctccggcccc gagtgcttca 180 caqccaacgg tgcagattac aggggaacac agagctggac agcgctgcaa ggtgggaagc 240 catgtctgtt ctggaacgag actttccagc atccgtacaa cacgctgaag taccccaacg 300 gggaaggagg actgggcgag cacaattatt gcagaaatcc agatggagac gtgagccctt 360 qqtqctacqt gqccqaqcat qaqqacqqaq tctactqqaa qtactqtqaa attcctqcct 420 qccaqatqcc tggaaacctt qqctqctaca aggatcatgg aaacccacct cctctcacqq 480 gcaccagtaa aacctctaac aagctcacca tacaaacctg tatcagcttc tgtcggagtc 540 agagattcaa gtttgctggg atggagtcag gctatgcctg cttctgtggg aacaatcctg 600 actactggaa gcacggggag gcggccagca ccgagtgcaa tagtgtctgc ttcggggacc 660 acacqcaqcc ctqcqqtqqq qacqqcaqqa ttatcctctt tqacactctc qtqqqcqcct 720 geggtgggaa ctacteagee atggeageeg tggtgtacte ceetgactte cetqacacet 780 acgccactgg cagagtetge tactggacca teegggttee aggageetet egeateeatt 840 tcaacttcac cctgtttgat atcagggact ctgcagacat ggtggagctg ctgqacqgct 900 acacceaccg cgtcctggtc cggctcagtg ggaggagccg cccgcctctg tctttcaatg 960 tetetetgga ttttgteatt ttgtatttet tetetgateg cateaateag geecagggat 1020 ttgctgtgtt gtaccaagcc accaaggagg aaccgccaca ggagagacct gctgtcaacc 1080 agaccetgge agaggtgate accgageaag ceaaceteag tgtcageget geccaetect 1140 ccaaaqteet etatqteate acceecaqee ccaqeeacee teegcagact geecaqqtaq 1200

77

| ccattcctgg | gcaccgtcag | ttagagccaa | cagccacaga | gtggaaggat | ggactgtgta | 1260 |
|------------|------------|------------|------------|------------|------------|------|
| | accetectea | | | | | |
| | aaatctcatc | | | | | |
| | ggagatatct | | | | | |
| | ctcaagggtc | | | | | |
| | ctgcatgaga | | | | | |
| | tccctgcctt | | | | | |
| | gcctgggatc | | | | | |
| | cccaggctga | | | | | |
| | atagtgtagg | | | | | |
| | cctgtcttta | | | | | |
| | cctggttccc | | | | | |
| | tggagctttc | | | | | |
| | gccacaccag | | | | | |
| | agectgagga | | | | | |
| | tagacttaac | | | | | |
| | gtcttcctgt | | | | | |
| tagagctggg | gctgtagcct | agagctgggg | ctgtagcaca | gagctggggc | tgtagcctag | 2280 |
| agctggggct | gtagcacaga | gctggggctg | tagcctagag | ctggggctgt | agcacagagc | 2340 |
| tggggctgta | gcacagagct | ggggctgtag | cctagagctg | gggctgtagc | acagagctgg | 2400 |
| ggctgtaact | cagcgatcaa | gagcttgctt | tgtatacatc | ggaccctagg | ttctatccca | 2460 |
| gcactatcag | aaggtgggag | agaaaaagac | tgcaccatag | catgcgggca | gcatctgtgg | 2520 |
| ttcctacgtg | aggtgtcatc | attttaaaag | cagatcaaaa | ctaccgcgag | ttttgtcctt | 2580 |
| tgtcccttat | catgggagca | gagtaggagt | aagggctctg | gtcttgctca | ttgtccccca | 2640 |
| gacagggagg | caggaaaagg | tcaggcttgg | gaactggaga | tcctcccagg | aaaagctgca | 2700 |
| agattgagag | acccagctgc | agttgggaga | ggaagggcca | tccccgactg | agaagtcctg | 2760 |
| cagtctggaa | gtggcctttg | tcagcagcag | ctgtgccctg | aaggtagacc | ttggtcactc | 2820 |
| | ccttgagcct | | | | | |
| | tctcagtcac | | | | | |
| | catctgggga | | | | | |
| | gtcaaccaga | | | | | |
| | gcctacaggg | | | | | |
| | ctcctggatg | | | | | |
| | gcaggttcca | | | | | |
| | gtcacagagt | | | | | |
| | atagcaggcc | | | | | |
| | cactcacctg | | | | | |
| | cggctcagca | | | | | |
| | ggcctggcac | | | | | |
| | aggaaaacta | | | | | |
| | acceaggeca | | | | | |
| | cacagagggc | | | | | |
| | agagcagccc | | | | | |
| | cacactgctt | | | | | |
| | caaggatagg | | | | | |
| | aatgtgggct | | | | | |
| | ggtacagctg | | | | | |
| ggtgccaccc | cacggctcac | tgtcatccca | ggagggaccc | cacctgatge | tcctcatcat | 4080 |

```
ccgctggcct gacactatca gagctcgcgc cggctgttgc cagggacaga ctgactacac 4140
ttgaccttca agagcactta gaagtggatg gcctccagac tctgtcagcc tctgcagggg 4200
ccacacaagt ctcccgagcc aagtccacaa gcctccatgg ttccctggct cctctcctgt 4260
ggagtgtcct gtttgatgtc tgaggtctgc tttgggtacc gccctgggaa ctgctaacct 4320
ccgattggtc cctttgtgtc tctgtttact gtcctcttct acctccaggt cacttagetc 4380
tggctgctct ggctgggagt ggggggtggg gatgctggct gcacccccac cctggtctgc 4440
caacagaacc tgggggcctc acacgggctc ctgtcttgcc aagctggagc tgagcacact 4500
ggcccaggct gagtggggca gagcaaacaa gtggaagggg atctctctcc ttagagggag 4560
gtggccgaag gtgtagatcc agcgagggag ctgccatccc cgccaccttc atagcagcaa 4620
gacettecca tttccaatet caccetecag cagggatatg actttggaca acaaggettt 4680
atttgtaaat atgctcttaa tatgcaactt tgagaataag atagaaacat catgtatttt 4740
aaaatataaa atgaagtgtg acacactgta tacaatttaa tatatttt taggattttg 4800
ttatttaaga aaatggaatg tgatggtact taacttttac aaaagagaga aaatgttatt 4860
4928
geggeege
```

<210> 68 <211> 1410 <212> DNA <213> Mus sp.

<400> 68

atagegeege eegeegeeg tetegegetg eteteegeeg etgegeteae tetggeggee 60 eggeeggege eggteeeeg eteeggeeee gagtgettea eageeaaegg tgeagattae 120 aggggaacac agagctggac agcgctgcaa ggtgggaagc catgtctgtt ctggaacgag 180 actttccaqc atccqtacaa cacqctgaag taccccaacg gggaaggagg actgggcgag 240 cacaattatt gcagaaatcc agatggagac gtgagccctt ggtgctacgt ggccgagcat 300 gaggacggag tetaetggaa gtaetgtgaa atteetgeet geeagatgee tggaaacett 360 ggctgctaca aggatcatgg aaacccacct cctctcacgg gcaccagtaa aacctctaac 420 aagctcacca tacaaacctg tatcagcttc tgtcggagtc agagattcaa gtttgctggg 480 atggaqtcag gctatgcctg cttctgtggg aacaatcctg actactggaa gcacggggag 540 qcqqccaqca ccqagtqcaa tagtqtctqc ttcggggacc acacqcaqcc ctgcggtggg 600 gacggcagga ttatcctctt tgacactctc gtgggcgcct gcggtgggaa ctactcagcc 660 atggcagecq tqqtqtactc ccctqacttc cctgacacct acgccactgg cagagtctgc 720 tactggacca teegggttee aggageetet egeateeatt teaactteae eetgtttgat 780 atcagggact ctgcagacat ggtggagctg ctggacggct acacccaccg cgtcctggtc 840 cggctcagtg ggaggagccg cccgcctctg tctttcaatg tctctctgga ttttgtcatt 900 ttgtatttct tctctgatcg catcaatcag gcccagggat ttgctgtgtt gtaccaagcc 960 accaaggagg aaccgccaca ggagagacct gctgtcaacc agaccctggc agaggtgatc 1020 accqaqcaag ccaacctcag tgtcagcgct gcccactcct ccaaagtcct ctatgtcatc 1080 acceccagee ceagecacee teegeagact geecaggtag ceatteetgg geacegteag 1140 ttggggccaa cagccacaga gtggaaggat ggactgtgta cggcctggcg accctcctca 1200 tcctcacagt cacagcagtt gtcgcaaaga ttcttctgca tgtcacattt aaatctcatc 1260 gagtecetge atcaggagae ettagggaet gtegteagee tgggggettet ggagatatet 1320 ggaccatttt ctatgaacct tccactacaa tctccatctt taagaagaag ctcaagggtc 1380 1410 agagtcaaca agatgaccgc aatcccctcg

<210> 69

<211> 470

<212> PRT

<213> Mus sp.

<400> 69

Met Ala Pro Pro Ala Ala Arg Leu Ala Leu Leu Ser Ala Ala Ala Leu

1 5 10 15

Thr Leu Ala Ala Arg Pro Ala Pro Gly Pro Arg Ser Gly Pro Glu Cys
20 25 30

Phe Thr Ala Asn Gly Ala Asp Tyr Arg Gly Thr Gln Ser Trp Thr Ala 35 40 45

Leu Gln Gly Gly Lys Pro Cys Leu Phe Trp Asn Glu Thr Phe Gln His 50 55 60

Pro Tyr Asn Thr Leu Lys Tyr Pro Asn Gly Glu Gly Gly Leu Gly Glu 65 70 75 80

His Asn Tyr Cys Arg Asn Pro Asp Gly Asp Val Ser Pro Trp Cys Tyr 85 90 95

Val Ala Glu His Glu Asp Gly Val Tyr Trp Lys Tyr Cys Glu Ile Pro 100 105 110

Ala Cys Gln Met Pro Gly Asn Leu Gly Cys Tyr Lys Asp His Gly Asn 115 120 125

Pro Pro Pro Leu Thr Gly Thr Ser Lys Thr Ser Asn Lys Leu Thr Ile 130 135 140

Gln Thr Cys Ile Ser Phe Cys Arg Ser Gln Arg Phe Lys Phe Ala Gly
145 150 155 160

Met Glu Ser Gly Tyr Ala Cys Phe Cys Gly Asn Asn Pro Asp Tyr Trp 165 170 175

Lys His Gly Glu Ala Ala Ser Thr Glu Cys Asn Ser Val Cys Phe Gly 180 185 190

Asp His Thr Gln Pro Cys Gly Gly Asp Gly Arg Ile Ile Leu Phe Asp 195 200 205

Thr Leu Val Gly Ala Cys Gly Gly Asn Tyr Ser Ala Met Ala Ala Val 210 215 220

| Val 225 | Tyr | Ser | Pro | Asp | Phe 230 | Pro | Asp | Thr | Tyr | Ala 235 | Thr | Gly | Arg | Val | Cys 240 |
|------------|------------|------------|------------|------------|------------|------------|-------------------|--------------------|------------|------------|------------|------------|-------------------|------------|------------|
| Tyr | Trp | Thr | Ile | Arg 245 | Val | Pro | Gly | Ala | Ser 250 | Arg | Ile | His | Phe | Asn 255 | Phe |
| Thr | Leu | Phe | Asp 260 | Ile | Arg | Asp | Ser | Ala 265 | Asp | Met | Val | Glu | Leu 270 | Leu | Asp |
| Gly | Tyr | Thr 275 | His | Arg | Val | Leu | Val 280 | Arg | Leu | Ser | Gly | Arg 285 | Ser | Arg | Pro |
| Pro | Leu 290 | Ser | Phe | Asn | Val | Ser 295 | Leu | Asp _. | Phe | Val | Ile 300 | Leu | Tyr | Phe | Phe |
| Ser 305 | Asp | Arg | Ile | Asn | Gln 310 | Ala | Gln | Gly | Phe | Ala 315 | Val | Leu | Tyr | Gln | Ala 320 |
| Thr | Lys | Glu | Glu | Pro 325 | Pro | Gln | Glu | Arg | Pro 330 | Ala | Val | Asn | Gln | Thr 335 | Leu |
| Ala | Glu | Val | Ile 340 | Thr | Glu | Gln | Ala | Asn 345 | Leu | Ser | Val | Ser | Ala 350 | Ala | His |
| Ser | Ser | Lys 355 | Val | Leu | Tyr | Val | 11e 360 | Thr | Pro | Ser | Pro | Ser 365 | His | Pro | Pro |
| Gln | Thr 370 | Ala | Gln | Val | Ala | Ile 375 | Pro | Gly | His | Arg | Gln 380 | Leu | Gly | Pro | Thr |
| Ala 385 | Thr | Glu | Trp | Lys | Asp 390 | Gly | Leu | Cys | Thr | Ala 395 | Trp | Arg | Pro | Ser | Ser 400 |
| Ser | Ser | Gln | Ser | Gln 405 | Gln | Leu | Ser | Gln | Arg 410 | Phe | Phe | Cys | Met | Ser 415 | |
| Leu | Asn | Leu | Ile 420 | Glu | Ser | Leu | His | Gln 4 25 | Glu | Thr | Leu | Gly | Thr 430 | Val | Val |
| Ser | Leu | Gly 435 | Leu | Leu | Glu | Ile | Ser 440 | Gly ' | Pro | Phe | Ser | Met 445 | Asn | Leu | Pro |
| Leu | Gln 450 | Ser | Pro | Ser | Leu | Arg 455 | Arg | Ser | Ser | Arg | Val 460 | Arg | Val | Asn | Lys |
| Met 465 | Thr | Ala | Ile | Pro | Ser 470 | | | | | | | | | | |

<210> 70 <211> 760

<212> PRT

<213> Mus sp.

<400> 70

Met Ala Leu Pro Ser Leu Gly Gln Asp Ser Trp Ser Leu Leu Arg Val 1 5 10 15

Phe Phe Phe Gln Leu Phe Leu Leu Pro Ser Leu Pro Pro Ala Ser Gly
20 25 30

Thr Gly Gly Gln Gly Pro Met Pro Arg Val Lys Tyr His Ala Gly Asp
35 40 45

Gly His Arg Ala Leu Ser Phe Phe Gln Gln Lys Gly Leu Arg Asp Phe
50 55 60

Asp Thr Leu Leu Ser Asp Asp Gly Asn Thr Leu Tyr Val Gly Ala 65 70 75 80

Arg Glu Thr Val Leu Ala Leu Asn Ile Gln Asn Pro Gly Ile Pro Arg 85 90 95

Leu Lys Asn Met Ile Pro Trp Pro Ala Ser Glu Arg Lys Lys Thr Glu 100 105 110

Cys Ala Phe Lys Lys Ser Asn Glu Thr Gln Cys Phe Asn Phe Ile 115 120 125

Arg Val Leu Val Ser Tyr Asn Ala Thr His Leu Tyr Ala Cys Gly Thr 130 135 140

Phe Ala Phe Ser Pro Ala Cys Thr Phe Ile Glu Leu Gln Asp Ser Leu 145 150 155 160

Leu Leu Pro Ile Leu Ile Asp Lys Val Met Asp Gly Lys Gly Gln Ser 165 170 175

Pro Leu Thr Leu Phe Thr Ser Thr Gln Ala Val Leu Val Asp Gly Met 180 185 190

Leu Tyr Ser Gly Thr Met Asn Asn Phe Leu Gly Ser Glu Pro Ile Leu 195 200 205

Met Arg Thr Leu Gly Ser His Pro Val Leu Lys Thr Asp Ile Phe Leu

| | 210 | | | | | 215 | | | | | 220 | | | | | |
|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|--|
| Arg 225 | Trp | Leu | His | Ala | Asp 230 | Ala. | Ser | Phe | Val | Ala 235 | Ala | Ile | Pro | Ser | Thr 240 | |
| Gln | Val | Val | Tyr | Phe 245 | Phe | Phe | Glu | Glu | Thr 250 | Ala | Ser | Glu | Phe | Asp 255 | Phe | |
| Phe | Glu | Glu | Leu 260 | Tyr | Ile | Ser | Arg | Val 265 | Ala | Gln | Val | Суз | Lys 270 | Asn | Asp | |
| Val | Gly | Gly 275 | Glu | Lys | Leu | Leu | Gln 280 | Lys | Lys | Trp | Thr | Thr 285 | Phe | Leu | Lys | |
| Ala | Gln 290 | Leu | Leu | Cys | Ala | G1n 295 | Pro | Gly | Gln | Leu | Pro 300 | Phe | Asn | Ile | Ile | |
| 305 | | | | | 310 | | | | | 315 | | | | Arg | 320 | |
| | | | | 325 | | | - | | 330 | | | | | Ser 335 | | |
| | | | 340 | | | | | 345 | | | | | 350 | Lys | | |
| - | - | 355 | | | | | 360 | | | | | 365 | | Tyr | - | |
| - | 370 | | | | | 375 | | Ī | | | 380 | | | Pro | | |
| 385 | • | | | | 390 | | | - | • | 395 | | | | Asp | 400 | |
| | | | | 405 | | | | | 410 | | | | | Tyr 415 | • | |
| | | | 420 | | | | | 425 | | - | | | 430 | His | | |
| | | 435 | | | | | 440 | | | | | 445 | | Val | | |
| | 450 | · | | | | 455 | | | | | 460 | | | Ser | | |
| Asp | Ser | Glu | Pro | Val | Arg | Asn | Leu | Gln | Leu | Ala | Pro | Ala | Gln | Gly | Ala | |

| 465 | | | | | 470 | | | | | 475 | | | | | 480 |
|------------|------------|------------|------------|-------------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|
| Val | Phe | Ala | Gly | Phe 485 | Ser | Gly | Gly | Ile | Trp 490 | Arg | Val | Pro | Arg | Ala 495 | Asn |
| Суз | Ser | Val | Tyr 500 | G1u | Ser | Суз | Val | Asp 505 | Суз | Val | Leu | Ala | Arg 510 | Азр | Pro |
| His | Суз | Ala 515 | Trp | Asp | Pro | Glu | Ser 520 | Arg | Leu | Cys | Ser | Leu 525 | Leu | Ser | Gly |
| Ser | Thr 530 | Lys | Pro | Trp | Lys | Gln 535 | Asp | Met | Glu | Arg | Gly 540 | Asn | Pro | Glu | Trp |
| Val 545 | Суз | Thr | Arg | Gly | Pro 550 | Met | Ala | Arg | Ser | Pro 555 | Arg | Arg | Gln | Ser | Pro 560 |
| Pro | Gln | Leu | Ile | Lys 565 | Glu | Val | Leu | Thr | Val 570 | Pro | Asn | Ser | Ile | Leu 575 | Glu |
| Leu | Arg | Cys | Pro 580 | His | Leu | Ser | Ala | Leu 585 | Ala | Ser | Tyr | His | Trp 590 | Ser | His |
| Gly | Arg | Ala 595 | Lys | Ile | Ser | Glu | Ala 600 | Ser | Ala | Thr | Val | Tyr 605 | Asn | Gly | Ser |
| Leu | Leu 610 | Leu | Leu | Pro | Gln | Азр 615 | Gly | Val | Gly | Gly | Leu 620 | Tyr | Gln | Cys | Val |
| Ala 625 | Thr | Glu | Asn | Gly | Tyr 630 | Ser | Tyr | Pro | Val | Val 635 | Ser | Tyr | Trp | Val | Asp 640 |
| Ser | Gln | Asp | Gln | Pro 645 | Leu | Ala | Leu | qeA | Pro 650 | Glu | Leu | Ala | Gly | Val 655 | Pro |
| Arg | Glu | Arg | Val 660 | Gln | Val | Pro | Leu | Thr 665 | Arg | Val | Gly | Gly | Gly 670 | Ala | Ser |
| Met | Ala | Ala 675 | Gln | Arg | Ser | Tyr | Trp 680 | Pro | His | Phe | Leu | Ile 685 | Val | Thr | Val |
| Leu | Leu 690 | Ala | Ile | Val | Leu | Leu 695 | Gly | Val | Leu | Thr | Leu 700 | Leu | Leu | Ala | Ser |
| Pro 705 | Leu | Gly | Ala | Leu | Arg 710 | Ala | Arg | Gly | Lys | Val 715 | Gln | Gly | Суз | Gly | Met 720 |
| Leu | Pro | Pro | Ara | Glu | Lvs | Ala | Pro | Len | Ser | Arg | Asp | Gln | His | Leu | Gln |

725 730 735

Pro Ser Lys Asp His Arg Thr Ser Ala Ser Asp Val Asp Ala Asp Asn 740 745 750

Asn His Leu Gly Ala Glu Val Ala 755 760

<210> 71

<211> 3046

<212> DNA

<213> Mus sp.

<400> 71

ctcggacgcc tgggttaggg gtctgtactg ctggggaacc atctggtgac catctcaggc 60 tgaccatggc cctaccatcc ctgggccagg actcatggag tctcctgcgt gttttttct 120 tccaactctt cctgctgcca tcactgccac ctgcttctgg gactggtggt caggggccca 180 tgcccagagt caaataccat gctggagacg ggcacagggc cctcagcttc ttccaacaaa 240 aaggeeteeg agaetttgae aegetgetee tgagtgaega tggcaacaet etetatgtgg 300 qqqctcqaqa qaccqtcctq gccttqaata tccagaaccc aggaatccca aggctaaaga 360 acatgatacc ctggccagcc agtgagagaa aaaagaccga atgtgccttt aagaagaaga 420 gcaatgagac acagtgtttc aacttcattc gagtcctggt ctcttacaat gctactcacc 480 tctatgcctg tgggaccttt gccttcagcc ctgcctgtac cttcattgaa ctccaagatt 540 ccctcctgtt gcccatcttg atagacaagg tcatggacgg gaagggccaa agccctttga 600 ccctqttcac aagcacacaa gctgtcttgg tcgatgggat gctttattcc ggcaccatga 660 acaactteet gggcagegag eccateetga tgeggacaet gggateeeat eetgttetea 720 agactgacat cttcttacgc tggctgcacg cggatgcctc cttcgtggca gccattccat 780 ccacccaggt cgtctatttc ttctttgagg agacagccag cgagtttgac ttctttgaag 840 agctqtatat atccaqqqtq qctcaaqtct qcaaqaacqa cqtqqqcqqt gaaaagctqc 900 tgcagaagaa gtggaccacc ttcctcaaag cccagttgct ctgcgctcag ccagggcagc 960 tgccattcaa catcatccgc cacgcggtcc tgctgcccgc cgattctccc tctgtttccc 1020 gcatctacgc agtctttacc tcccagtggc aggttggcgg gaccaggagc tcagcagtct 1080 qtqccttctc tctcacggac attgagcgag tctttaaagg gaagtacaag gagctgaaca 1140 aggagacete cegetggace aettaceggg geteagaggt cageeegagg ceaggeagtt 1200 qctccatggg cccctcctct gacaaagcct tgaccttcat gaaggaccat tttctgatgg 1260 atgagcacgt ggtaggaaca cccctgctgg tgaagtctgg tgtggagtac acacggcttg 1320 ctgtggagtc agctcggggc cttgatggga gcagccatgt ggtcatgtat ctgggtacct 1380 ccacqqqtcc cctqcacaaq qctqtqqtqc ctcaqqacaq cagtqcttat ctcqtqqaqg 1440 agattcagct gagccctgac tctgagcctg ttcgaaacct gcagctggcc cccgcccagg 1500 qtqcaqtqtt tgcaggcttc tctggaggca tctggagagt tcccagggcc aattgcagtg 1560 tctacgagag ctgtgtggac tgtgtgcttg ccagggaccc tcactgtgcc tgggaccctg 1620 aatcaaqact ctgcagcctt ctgtctggct ctaccaagcc ttggaagcag gacatggaac 1680 qcqqcaaccc qqaqtqqqta tqcacccqtq qccccatqqc caggagcccc cggcgtcaga 1740 geocecetea actaattaaa gaagteetga cagteeceaa etecateetg gagetgeget 1800 geocecacet gloageactg geotettace actggagtea tggccgagec aaaateteag 1860 aageetetge tacegtetae aatggeteee tettgetget geegeaggat ggtgtegggg 1920 geetetacea gtgtgtgggg actgagaacg getacteata eeetgtggte teetattggg 1980

```
tagacageca ggaccagece etggegetgg accetgaget ggegggegtt eccegtgage 2040
gtgtgcaggt cccgctgacc agggtcggag gcggagcttc catggctgcc cagcggtcct 2100
actggcccca ttttctcatc gttaccgtcc tcctggccat cgtgctcctg ggagtgctca 2160
ctctcctcct cgcttcccca ctgggggcgc tgcgggctcg gggtaaggtt cagggctgtg 2220
ggatgctgcc ccccagggaa aaggctccac tgagcaggga ccagcacctc cagccctcca 2280
aggaccacag gacctctgcc agtgacgtag atgccgacaa caaccatctg ggcgccgaag 2340
atgccaggca cagtgccact ctgaccaggg taggaggctc tcctgctaac gtgtgtcacc 2460
tacagcaccc agtaggtcct cccctgtggg actctcttct gcaagcacat tgggctgtct 2520
ccatacctgt acttgtgctg tgacaggaag agccagacag gtttctttga ttttgattga 2580
cccaagagcc ctgcctgtaa caaacgtgct ccaggagacc atgaaaggtg tggctgtctg 2640
ggattctgtg gtgacaaacc taagcatccg agcaagctgg ggctattcct gcaaactcca 2700
tcctgaacgc tgtcactcta gaagcagctg ctgctttgaa caccagccca ccctccttcc 2760
caagagtete tatggagttg geceettgtg ttteetttae eagtegggee atactgtttg 2820
ggaagtcatc tctgaagtct aaccaccttc cttcttggtt cagtttggac agattgttat 2880
tattgtctct gccctggcta gaatgggggc ataatctgag ccttgttccc ttgtccagtg 2940
tggctgaccc ttgacctctt cettectect ceetttgttt tgggattcag aaaactgett 3000
                                                              3046
qtcacaqaca atttatttt tattaaaaaa gatataagct ttaaag
```

<210> 72 <211> 2915 <212> DNA <213> Mus sp.

<400> 72

gtcgacccac gcgtccggcc gcgcgtcctt ctgccggctt cagctcgtat ccccggagtc 60 caccegeceg teeeggggtg eggactggee etgagetgge egtaeagece ggetteggae 120 ggtcctcgct ggagccatgg gccgccggct cggcagggtg gcggcgctgc tgctcgggct 180 gctagtggag tgcactgagg ccaaaaaaca ttgctggtat tttgaaggac tctatcccac 240 atactatata tgccgttcct atgaagactg ctgtggctcc aggtgctgtg tgagggccct 300 ttccatacag aggctgtggt atttttggtt cctgctgatg atgggtgtgc tgttctgctg 360 tggtgccggt ttcttcattc gccggcgcat gtatccgcca ccactcattg aggagcccac 420 attcaatgtg tectatacea ggeageeace aaateetget eeaggageae ageaaatggg 480 accgccatat tacaccgacc ctggaggacc cgggatgaat cctgttggca ataccatggc 540 tatggctttc caggtccagc ccaattcacc tcacggaggc acaacttacc caccccctcc 600 ttcctactgc aacacgcctc cacccccta tgaacaggtg gtgaaggaca agtagcaaga 660 tgctacatca aaggcaaaga ggatggacag gcccttttgt ttaccttccc atcctcaccg 720 atacttgctg atagggtggt ccaagggaaa acttggatat tctcaaagca agcccagctc 780 tettteaagt ettttgtgga ggacatttga atecacactg teteetetgt tgettetgtt 840 tetgatgtag tetgtgetet etgagagagt gtggcaacag teeetgaggg ttgatattee 900 tagggtgtcc agggtagatc ctcgggagag aggctaaggg gaaaggaagg catagcctgt 960 gtgttagggg gcagataaag tggtcaggct gagataagac tcacatgatg cagtagttgg 1020 cagtgaactt cgaagagaca ctatccacca tcccagccca ttctcctaat agaagctgtg 1080 gggctgtgtt gttgatgctc tttggtctcc actcacattt tgaaaatagg ctttcctctg 1140 caggaatagg aaagacccaa gtacatattt gcttccactt aaaaatgagg gtcagaacca 1200 ggcctcagtt ggacatctat agttaaataa aggccattag agaggggaaa tctttaagtt 1260

```
agtgcatgta ttgaagtgag ggtgtccttt gagatcagat ggggagagtg aactctgcqq 1380
ggggtggggt gtctctactc agagggctcc aacacccttt tcttaggtag ttctggtqat 1440
gggttttatg ggcactatag agctgagggg cacattaggc cgggtagtta cattgaccct 1500
agctagggtt gtatgtgatc ccaacagaga tgtgctggcc tcagaaqagg ggacgtttgt 1620
qqataqaqcc qtqaaaacct acttaqttqc acagatgaca taatcaaaag tagagaaaga 1680
aqtqtaqtta qagatqccat ttcccagqtg agaatcagag ctcatccata gatttacaag 1740
tagtggctgg agttaacagt atggagttct tttcccttgc gtagttagtc acgttgatgt 1800
gtatttaaac ccaggttgag accttgtgta ctaagagcaa ggaagtatag ctaagatgtc 1860
tagattattt atatgtagta tggtggggag tggggctgca aggaaggggg ctgacattgt 1920
aaatgagaaa atcagagcca tttgataaac tgttacttgt tggatcaggc atccaaaagt 1980
gtctcttgag tggacattga gtattcttta ccacctacaa gaccaggagg catggtgtca 2040
ttctccattg gggtatttat atgaggtaga ggttcaggaa tcgacagtag ctgtgtgggc 2100
ttaqtttaag gactgaaagc atagggactg gtagacagtt tcataggaaa ctgcggggaa 2160
ggaatggata cctttaaaga cagtttgtgg atgcagatgc tgccacccat cattgagcac 2220
cettgtgtet etggetteet gteaetggat eeagtacece teeatgettg ggteettqtt 2280
ttacataaga caacaaagca caatgtctqc tqtttacaat caagacgact acatggtcca 2340
aacatttett etetetteta teaettgtgg etttaaette eattteetee gtteettttt 2400
aaaatcaaga agcacagtca gagctgcccc tgggattgca tcagggaacg gctgatcaag 2460
qcattcagtg tccatgacta aatcttatct ttttgatagc aaatcctttt aagaaactga 2520
acaattgcta aggctcagca attttatact ccaatgtctg tgtaaggtaa attttgtttg 2580
ccattgagec cacattggaa tteettetga egteaacaet gacaatgeet atggaaattg 2640
cacttetggg tatatgteec ageateettg ttttettatg tttggtgagt aaggeteace 2700
ccttccagca gctctacttc tgtgtgctga ggtcctgtag agccggggct tgggcacaga 2760
catgaggcag acttgtgcat gctctttctt ggcaacactt ggctcatatt tcttgttctc 2820
ttttgataga gtcctgtttc ctatgtattt aaaaaataat aaaagtgaat ttagtcaaaa 2880
aaaaaaaaaa aaaaaaaaaa aaaaagggcg gccgc
                                                               2915
```

<210> 73 <211> 516 <212> DNA <213> Mus sp.

<400> 73

atgggccgcc ggctcggcag ggtggcggc ctgctgctcg ggctgctagt ggagtgcact 60 gaggccaaaa aacattgctg gtattttgaa ggactctatc ccacatacta tatatgccgt 120 tcctatgaag actgctgtg ctccaggtgc tgtgtgaggg ccctttccat acagaggctg 180 tggtattttt ggttcctgct gatgatgggt gtgctgttct gctgtggtgc cggtttcttc 240 attcgccggc gcatgtatcc gccaccactc attgaggagc ccacattcaa tgtgtcctat 300 accaggcagc caccaaatcc tgctccagga gcacagcaaa tgggaccgcc atattacacc 360 gaccctggag gacccgggat gaatcctgtt ggcaatacca tggctatggc tttccaggtc 420 cagcccaatt cacctcacgg aggcacaact tacccaccc ctccttccta ctgcaacacg 480 cctccaccc cctatgaaca ggtggtgaag gacaag

<210> 74 <211> 172

<212> PRT

<213> Mus sp.

<400> 74

Met Gly Arg Arg Leu Gly Arg Val Ala Ala Leu Leu Leu Gly Leu Leu 1 5 10 15

Val Glu Cys Thr Glu Ala Lys Lys His Cys Trp Tyr Phe Glu Gly Leu
20 25 30

Tyr Pro Thr Tyr Tyr Ile Cys Arg Ser Tyr Glu Asp Cys Cys Gly Ser
35 40 45

Arg Cys Cys Val Arg Ala Leu Ser Ile Gln Arg Leu Trp Tyr Phe Trp 50 55 60

Phe Leu Leu Met Met Gly Val Leu Phe Cys Cys Gly Ala Gly Phe Phe 65 70 75 80

Ile Arg Arg Arg Met Tyr Pro Pro Pro Leu Ile Glu Glu Pro Thr Phe
85 90 95

Asn Val Ser Tyr Thr Arg Gln Pro Pro Asn Pro Ala Pro Gly Ala Gln 100 105 110

Gln Met Gly Pro Pro Tyr Tyr Thr Asp Pro Gly Gly Pro Gly Met Asn 115 120 125

Pro Val Gly Asn Thr Met Ala Met Ala Phe Gln Val Gln Pro Asn Ser 130 135 140

Pro His Gly Gly Thr Thr Tyr Pro Pro Pro Pro Ser Tyr Cys Asn Thr 145 150 155 160

Pro Pro Pro Pro Tyr Glu Gln Val Val Lys Asp Lys 165 170

<210> 75

<211> 398

<212> PRT

<213> Homo sapiens

<400> 75

Met Trp Leu Leu Thr Met Ala Ser Leu Ile Ser Val Leu Gly Thr

1 5 10 15

Thr His Gly Leu Phe Gly Lys Leu His Pro Gly Ser Pro Glu Val Thr

| | | | 20 | | | | | 25 | | | | | 30 | | |
|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|-------------|------------|
| Met | Asn | Ile 35 | Ser | Gln | Met | Ile | Thr 40 | Tyr | Trp | Gly | Tyr | Pro 45 | Asn | Gl u | Glu |
| Tyr | Glu 50 | Val | Val | Thr | Glu | Asp 55 | Gly | Tyr | Ile | Leu | Glu 60 | Val | Asn | Arg | Ile |
| Pro 65 | Tyr | Gly | Lys | Lys | Asn 70 | Ser | Gly | Asn | Thr | Gly 75 | Gln | Arg | Pro | Val | Val 80 |
| Phe | Leu | Gln | His | Gly 85 | Leu | Leu | Ala | Ser | Ala 90 | Thr | Asn | Trp | Ile | Ser 95 | Asn |
| Leu | Pro | Asn | Asn 100 | Ser | Leu | Ala | Phe | Ile 105 | Leu | Ala | Asp | Ala | Gly 110 | Tyr | Asp |
| Val | Trp | Leu 115 | Gly | Asn | Ser | Arg | Gly 120 | Asn | Thr | Trp | Ala | Arg 125 | Arg | Asn | Leu |
| Tyr | Tyr 130 | Ser | Pro | Asp | Ser | Val 135 | Glu | Phe | Trp | Ala | Phe 140 | Ser | Phe | Asp | Glu |
| Met 145 | Ala | Lys | Tyr | Asp | Leu 150 | Pro | Ala | Thr | Ile | Asp 155 | Phe | Ile | Val | Lys | Lys 160 |
| Thr | Gly | Gln | Lys | Gln 165 | Leu | His | Tyr | Val | Gly 170 | His | Ser | Gln | Gly | Thr 175 | Thr |
| Ile | Gly | Phe | Ile 180 | Ala | Phe | Ser | Thr | Asn 185 | Pro | Ser | Leu | Ala | Lys 190 | Arg | Ile |
| Lys | Thr | Phe 195 | Tyr | Ala | Leu | Ala | Pro 200 | Val | Ala | Thr | Val | Lys 205 | Tyr | Thr | Lys |
| Ser | Leu 210 | Ile | Asn | Lys | Leu | Arg 215 | Phe | Val | Pro | Gln | Ser 220 | Leu | Phe | Lys | Phe |
| Ile 225 | Phe | Gly | Asp | Lys | Ile 230 | Phe | Tyr | Pro | His | Asn 235 | Phe | Phe | Asp | Gln | Phe 240 |
| Leu | Ala | Thr | Glu | Val 245 | Cys | Ser | Arg | Glu | Met 250 | Leu | Asn | Leu | Leu | Суз 255 | Ser |
| Asn | Δ1 = | Len | Phe | Tle | Tle | Cue | G] v | Phe | Acr | Sar | T.ve | Δer | Dhe | Non- | Thr |

Ser Arg Leu Asp Val Tyr Leu Ser His Asn Pro Ala Gly Thr Ser Val

Gln Asn Met Phe His Trp Thr Gln Ala Val Lys Ser Gly Lys Phe Gln Ala Tyr Asp Trp Gly Ser Pro Val Gln Asn Arg Met His Tyr Asp Gln Ser Gln Pro Pro Tyr Tyr Asn Val Thr Ala Met Asn Val Pro Ile Ala Val Trp Asn Gly Gly Lys Asp Leu Leu Ala Asp Pro Gln Asp Val Gly

Leu Leu Pro Lys Leu Pro Asn Leu Ile Tyr His Lys Glu Ile Pro

Phe Tyr Asn His Leu Asp Phe Ile Trp Ala Met Asp Ala Pro Gln Glu

Val Tyr Asn Asp Ile Val Ser Met Ile Ser Glu Asp Lys Lys

<210> 76

<211> 760

<212> PRT

<213> Mus sp.

Met Ala Leu Pro Ser Leu Gly Gln Asp Ser Trp Ser Leu Leu Arg Val

Phe Phe Gln Leu Phe Leu Leu Pro Ser Leu Pro Pro Ala Ser Gly

Thr Gly Gly Gln Gly Pro Met Pro Arg Val Lys Tyr His Ala Gly Asp

Gly His Arg Ala Leu Ser Phe Phe Gln Gln Lys Gly Leu Arg Asp Phe

Asp Thr Leu Leu Ser Asp Asp Gly Asn Thr Leu Tyr Val Gly Ala

Arg Glu Thr Val Leu Ala Leu Asn Ile Gln Asn Pro Gly Ile Pro Arg

| Leu | Lys | Asn | Met 100 | lle | Pro | Trp | Pro | Ala 105 | Ser | Glu | Arg | Lys | Lys 110 | Thr | Glu |
|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|
| Cys | Ala | Phe 115 | Lys | Lys | Lys | Ser | Asn 120 | Glu | Thr | Gln | Суз | Phe 125 | Asn | Phe | Ile |
| Arg | Val 130 | Leu | Val | Ser | Tyr | Asn 135 | Ala | Thr | His | Leu | Tyr 140 | Ala | Cys | Gly | Thr |
| Phe 145 | Ala | Phe | Ser | Pro | Ala 150 | Cys | Thr | Phe | Ile | Glu 155 | Leu | Gln | Asp | Ser | Leu 160 |
| Leu | Leu | Pro | Ile | Leu 165 | Ile | Asp | Lуз | Val | Met 170 | Asp | Gly | Lys | Gly | Gln 175 | Ser |
| Pro | Leu | Thr | Leu 180 | Phe | Thr | Ser | Thr | Gln 185 | Ala | Val | Leu | Val | Asp 190 | Gly | Met |
| Leu | Tyr | Ser 195 | Gly | Thr | Met | Asn | Asn 200 | Phe | Leu | Gly | Ser | Glu 205 | Pro | Ile | Leu |
| Met | Arg 210 | Thr | Leu | Gly | Ser | His 215 | Pro | Val | Leu | Lys | Thr 220 | Asp | Ile | Phe | Leu |
| Arg 225 | Trp | Leu | His | Ala | Asp 230 | Ala | Ser | Phe | Val | Ala 235 | Ala | Ile | Pro | Ser | Thr 240 |
| Gln | Val | Val | Туг | Phe 245 | Phe | Phe | Glu | Glu | Thr 250 | Ala | Ser | Glu | Phe | Asp 255 | Phe |
| Phe | Glu | Glu | Leu 260 | Tyr | Ile | Ser | Arg | Val 265 | Ala | Gln | Val | Суз | Lys 270 | Asn | Asp |
| Val | - | Gly 275 | Glu | Lys | Leu | | Gln 280 | _ | Lys | Trp | Thr | Thr 285 | Phe | Leu | Lys |
| Ala | Gln 290 | Leu | Leu | Суз | Ala | Gln 295 | Pro | Gly | Gln | Leu | Pro 300 | Phe | Asn | Ile | Ile |
| Arg 305 | His | Ala | Val | Leu | Leu 310 | Pro | Ala | Asp | Ser | Pro 315 | Ser | Val | Ser | Arg | 11e 320 |
| Tyr | Ala | Val | Phe | Thr 325 | Ser | Gln | Trp | Gln | Val 330 | Gly | Gly | Thr | Arg | Ser 335 | Ser |
| Ala | Val | Cys | Ala 340 | Phe | Ser | Leu | Thr | Asp 345 | Ile | Glu | Arg | Val | Phe 350 | Lys | Gly |

| Lys | Tyr | Lys 355 | Glu | Leu | Asn | Lys | Glu 360 | Thr | Ser | Arg | Trp | Thr 365 | Thr | Tyr | Arg |
|------------|------------|------------|------------|-------------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|
| Gly | Ser 370 | Glu | Val | Ser | Pro | Arg 375 | Pro | Gly | Ser | Суз | Ser 380 | Met | Gly | Pro | Ser |
| Ser 385 | Asp | Lys | Ala | Leu | Thr 390 | Phe | Met | Lys | Asp | His 395 | Phe | Leu | Met | Asp | Glu 400 |
| His | Val | Val | Gly | Thr 405 | Pro | Leu | Leu | Val | Lys 410 | Ser | Gly | Val | Glu | Tyr 415 | Thr |
| Arg | Leu | Ala | Val 420 | Glu | Ser | Ala | Arg | Gly 425 | Leu | Asp | Gly | Ser | Ser 430 | His | Val |
| Val | Met | Tyr 435 | Leu | Gly | Thr | Ser | Thr 440 | Gly | Pro | Leu | His | Lys 445 | Ala | Val | Val |
| Pro | Gln 450 | Asp | Ser | Ser | Ala | Tyr 455 | Leu | Val | Glu | Glu | Ile 460 | Gln | Leu | Ser | Pro |
| Asp 465 | Ser | Glu | Pro | Val | Arg 470 | Asn | Leu | Gln | Leu | Ala 475 | Pro | Ala | Gln | Gly | Ala 480 |
| Val | Phe | Ala | Gly | Phe 485 | Ser | Gly | Gly | Ile | Trp 490 | Arg | Val | Pro | Arg | Ala 495 | Asn |
| Суѕ | Ser | Val | Tyr 500 | Glu | Ser | Суз | Val | Asp 505 | Суз | Val | Leu | Ala | Arg 510 | Asp | Pro |
| His | Суз | Ala 515 | Trp | Asp | Pro | Glu | Ser 520 | Arg | Leu | Суз | Ser | Leu 525 | Leu | Ser | Gly |
| Ser | Thr 530 | Lys | Pro | Trp | Lys | Gln 535 | _ | Met | Glu | Arg | Gly 540 | Asn | Pro | Glu | Trp |
| Val 545 | Суз | Thr | Arg | Gly | Pro 550 | Met | Ala | Arg | Ser | Pro 555 | Arg | Arg | Gln | Ser | Pro 560 |
| Pro | Gln | Leu | Ile | Lys 565 | Glu | Val | Leu | Thr | Val 570 | Pro | Äsn | Ser | Ile | Leu 575 | Glu |
| Leu | Arg | Суз | Pro 580 | His | Leu | Ser | Ala | Leu 585 | Ala | Ser | Tyr | His | Trp 590 | Ser | His |
| Gly | Arg | Ala 595 | Ьys | Ile | Ser | Glu | Ala 600 | Ser | Ala | Thr | Val | Tyr 605 | Asn | Gly | Ser |

Leu Leu Leu Pro Gln Asp Gly Val Gly Gly Leu Tyr Gln Cys Val 620 615 610 Ala Thr Glu Asn Gly Tyr Ser Tyr Pro Val Val Ser Tyr Trp Val Asp 635 630 625 Ser Gln Asp Gln Pro Leu Ala Leu Asp Pro Glu Leu Ala Gly Val Pro 650 645 Arg Glu Arg Val Gln Val Pro Leu Thr Arg Val Gly Gly Ala Ser 665 Met Ala Ala Gln Arg Ser Tyr Trp Pro His Phe Leu Ile Val Thr Val 680 Leu Leu Ala Ile Val Leu Leu Gly Val Leu Thr Leu Leu Leu Ala Ser 700 690 695 Pro Leu Gly Ala Leu Arg Ala Arg Gly Lys Val Gln Gly Cys Gly Met 710 715 705 Leu Pro Pro Arg Glu Lys Ala Pro Leu Ser Arg Asp Gln His Leu Gln 725 730 Pro Ser Lys Asp His Arg Thr Ser Ala Ser Asp Val Asp Ala Asp Asn 745 750 740 Asn His Leu Gly Ala Glu Val Ala 760 755 <210> 77 <211> 3046 <212> DNA <213> Mus sp. <400> 77 ctcggacgcc tgggttaggg gtctgtactg ctggggaacc atctggtgac catctcaggc 60 tgaccatggc cctaccatcc ctgggccagg actcatggag tctcctgcgt gttttttct 120 tccaactctt cctgctgcca tcactgccac ctgcttctgg gactggtggt caggggccca 180 tgcccagagt caaataccat gctggagacg ggcacagggc cctcagcttc ttccaacaaa 240 aaggeeteeg agaetttgae aegetgetee tgagtgaega tggeaacaet etetatgtgg 300 gggctcgaga gaccgtcctg gccttgaata tccagaaccc aggaatccca aggctaaaga 360 acatgatacc ctggccagcc agtgagagaa aaaagaccga atgtgccttt aagaagaaga 420 gcaatgagac acagtgtttc aacttcattc gagtcctggt ctcttacaat gctactcacc 480

tetatgeetg tgggacettt geetteagee etgeetgtae etteattgaa eteeaagatt 540 eeeteetgtt geeeatettg atagacaagg teatggacgg gaagggeeaa ageeetttga 600 eeetgtteae aageacacaa getgtettgg tegatgggat getttattee ggeaceatga 660

```
acaactteet gggeagegag eccateetga tgeggaeaet gggateeeat eetgttetea 720
agactgacat cttcttacgc tggctgcacg cggatgcctc cttcgtggca gccattccat 780
ccacccaggt cgtctatttc ttctttgagg agacagccag cgagtttgac ttctttgaag 840
agctgtatat atccagggtg gctcaagtct gcaagaacga cgtgggcggt gaaaagctgc 900
tgcagaagaa gtggaccacc ttcctcaaag cccagttgct ctgcgctcag ccagggcagc 960
tgccattcaa catcatccgc cacgcggtcc tgctgcccgc cgattctccc tctgtttccc 1020
gcatctacgc agtctttacc tcccagtggc aggttggcgg gaccaggagc tcagcagtct 1080
gtgccttctc tctcacggac attgagcgag tctttaaagg gaagtacaag gagctgaaca 1140
aggagacete eegetggace acttaceggg geteagaggt eageeegagg eeaggeagtt 1200
gctccatggg cccctcctct gacaaagcct tgaccttcat gaaggaccat tttctgatgg 1260
atgagcacgt ggtaggaaca cccctgctgg tgaagtctgg tgtggagtac acacggcttg 1320
ctgtggagtc agctcggggc cttgatggga gcagccatgt ggtcatgtat ctgggtacct 1380
ccacgggtcc cctgcacaag gctgtggtgc ctcaggacag cagtgcttat ctcgtggagg 1440
agattcagct gagccctgac tctgagcctg ttcgaaacct gcagctggcc cccgcccagg 1500
gtgcagtgtt tgcaggcttc tctggaggca tctggagagt tcccagggcc aattgcagtg 1560
tctacgagag ctgtgtggac tgtgtgcttg ccagggaccc tcactgtgcc tgggaccctg 1620
aatcaagact ctgcagcctt ctgtctggct ctaccaagcc ttggaagcag gacatggaac 1680
geggeaacce ggagtgggta tgcaccegtg geeccatgge caggageece eggegteaga 1740
geoccectea actaattaaa gaagteetga eagteeceaa etecateetg gagetgeget 1800
gececeacet gteageactg geetettace actggagtea tggccgagee aaaateteag 1860
aagcetetge tacegtetae aatggeteee tettgetget geegeaggat ggtgtegggg 1920
gcctctacca gtgtgtggcg actgagaacg gctactcata ccctgtggtc tcctattggg 1980
tagacageca ggaccagece etggegetgg accetgaget ggegggegtt eccegtgage 2040
gtgtgcaggt cccgctgacc agggtcggag gcggagcttc catggctgcc cagcggtcct 2100
actggcccca ttttctcatc gttaccgtcc tcctggccat cgtgctcctg ggagtgctca 2160
ctctcctcct cgcttcccca ctgggggcgc tgcgggctcg gggtaaggtt cagggctgtg 2220
ggatgctgcc ccccagggaa aaggctccac tgagcaggga ccagcacctc cagccctcca 2280
aggaccacag gacctctgcc agtgacgtag atgccgacaa caaccatctg ggcgccgaag 2340
atgccaggca cagtgccact ctgaccaggg taggaggctc tcctgctaac gtgtgtcacc 2460
tacagcaccc agtaggtcct cccctgtggg actctcttct gcaagcacat tgggctgtct 2520
ccatacctgt acttgtgctg tgacaggaag agccagacag gtttctttga ttttgattga 2580
cccaagagcc ctgcctgtaa caaacgtgct ccaggagacc atgaaaggtg tggctgtctg 2640
ggattctgtg gtgacaaacc taagcatccg agcaagctgg ggctattcct gcaaactcca 2700
tcctgaacgc tgtcactcta gaagcagctg ctgctttgaa caccagccca ccctccttcc 2760
caagagtoto tatggagttg goccottgtg tttootttac cagtogggoo atactgtttg 2820
ggaagtcate tetgaagtet aaccacette ettettggtt cagtttggae agattgttat 2880
tattgtctct gccctggcta gaatgggggc ataatctgag ccttgttccc ttgtccagtg 2940
tggctgaccc ttgacctctt ccttcctcct ccctttgttt tgggattcag aaaactgctt 3000
gtcacagaca atttatttt tattaaaaaa gatataagct ttaaag
                                                                 3046
```

<210> 78 <211> 1436

<212> PRT

<213> Bos sp.

<400> 78

| Met 1 | Ala | Leu | Gly | Arg 5 | His | Leu | Ser | Leu | Arg 10 | Gly | Leu | Cys | Val | Leu 15 | Leu |
|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|
| Leu | Gly | Thr | Met 20 | Val | Gly | Gly | Gln | Ala 25 | Leu | Glu | Leu | Arg | Leu 30 | Lys | Asp |
| Gly | Val | His 35 | Arg | Cys | Glu | Gly | Arg 40 | Val | Glu | Val | Lys | His 45 | Gln | Gly | Glu |
| Trp | Gly 50 | Thr | Val | Asp | Gly | Tyr 55 | Arg | Trp | Thr | Leu | Lỳs 60 | Asp | Ala | Ser | Val |
| Val 65 | Cys | Arg | Gln | Leu | Gly 70 | Cys | Gly | Ala | Ala | Ile 75 | Gly | Phe | Pro | Gly | Gly 80 |
| Ala | Tyr | Phe | Gly | Pro 85 | Gly | Leu | Gly | Pro | Ile 90 | Trp | Leu | Leu | Tyr | Thr 95 | Ser |
| Суз | Glu | Gly | Thr 100 | Glu | Ser | Thr | Val | Ser 105 | Asp | Суз | Glu | His | Ser 110 | Asn | Ile |
| Lys | Asp | Tyr 115 | Arg | Asn | Asp | Gly | Tyr 120 | Asn | His | Gly | Arg | Asp 125 | Ala | Gly | Val |
| Val | Cys 130 | Ser | Gly | Phe | Val | Arg 135 | Leu | Ala | Gly | Gly | Asp 140 | Gly | Pro | Cys | Ser |
| Gly 145 | Arg | Val | Glu | Val | His 150 | Ser | Gly | Glu | Ala | Trp 155 | Ile | Pro | Val | Ser | Asp 160 |
| Gly | Asn | Phe | Thr | Leu 165 | Ala | Thr | Ala | Gln | Ile 170 | Ile | Cys | Ala | Glu | Leu 175 | Gly |
| Cys | Gly | Lys | Ala 180 | Val | Ser | Val | Leu | Gly 185 | His | Glu | Leu | Phe | Arg 190 | Glu | Ser |
| Ser | Ala | Gln 195 | Val | Trp | Ala | Glu | Glu 200 | Phe | Arg | Суз | Glu | Gly 205 | Glu | Glu | Pro |
| Glu | Leu 210 | Trp | Val | Суз | Pro | Arg 215 | Val | Pro | Суз | Pro | Gly 220 | Gly | Thr | Cys | His |
| His 225 | Ser | Gly | Ser | Ala | Gln 230 | Val | Val | Cys | Ser | Ala 235 | Tyr | Ser | Glu | Val | Arg 240 |
| Leu | Met | Thr | Asn | Gly 245 | Ser | Ser | Gln | Cys | G1u 250 | Gly | Gln. | Val | Glu | Met 255 | Asn |

Ile Ser Gly Gln Trp Arg Ala Leu Cys Ala Ser His Trp Ser Leu Ala Asn Ala Asn Val Ile Cys Arg Gln Leu Gly Cys Gly Val Ala Ile Ser Thr Pro Gly Gly Pro His Leu Val Glu Glu Gly Asp Gln Ile Leu Thr Ala Arg Phe His Cys Ser Gly Ala Glu Ser Phe Leu Trp Ser Cys Pro Val Thr Ala Leu Gly Gly Pro Asp Cys Ser His Gly Asn Thr Ala Ser Val Ile Cys Ser Gly Asn Gln Ile Gln Val Leu Pro Gln Cys Asn Asp Ser Val Ser Gln Pro Thr Gly Ser Ala Ala Ser Glu Asp Ser Ala Pro Tyr Cys Ser Asp Ser Arg Gln Leu Arg Leu Val Asp Gly Gly Pro Cys Ala Gly Arg Val Glu Ile Leu Asp Gln Gly Ser Trp Gly Thr Ile Cys Asp Asp Gly Trp Asp Leu Asp Asp Ala Arg Val Val Cys Arg Gln Leu Gly Cys Gly Glu Ala Leu Asn Ala Thr Gly Ser Ala His Phe Gly Ala Gly Ser Gly Pro Ile Trp Leu Asp Asn Leu Asn Cys Thr Gly Lys Glu Ser His Val Trp Arg Cys Pro Ser Arg Gly Trp Gly Gln His Asn Cys Arg His Lys Gln Asp Ala Gly Val Ile Cys Ser Glu Phe Leu Ala Leu Arg Met Val Ser Glu Asp Gln Gln Cys Ala Gly Trp Leu Glu Val

Phe Tyr Asn Gly Thr Trp Gly Ser Val Cys Arg Asn Pro Met Glu Asp

Ile Thr Val Ser Thr Ile Cys Arg Gln Leu Gly Cys Gly Asp Ser Gly Thr Leu Asn Ser Ser Val Ala Leu Arg Glu Gly Phe Arg Pro Gln Trp Val Asp Arg Ile Gln Cys Arg Lys Thr Asp Thr Ser Leu Trp Gln Cys Pro Ser Asp Pro Trp Asn Tyr Asn Ser Cys Ser Pro Lys Glu Glu Ala Tyr Ile Trp Cys Ala Asp Ser Arg Gln Ile Arg Leu Val Asp Gly Gly Gly Arg Cys Ser Gly Arg Val Glu Ile Leu Asp Gln Gly Ser Trp Gly Thr Ile Cys Asp Asp Arg Trp Asp Leu Asp Asp Ala Arg Val Val Cys. Lys Gln Leu Gly Cys Gly Glu Ala Leu Asp Ala Thr Val Ser Ser Phe Phe Gly Thr Gly Ser Gly Pro Ile Trp Leu Asp Glu Val Asn Cys Arg Gly Glu Glu Ser Gln Val Trp Arg Cys Pro Ser Trp Gly Trp Arg Gln His Asn Cys Asn His Gln Glu Asp Ala Gly Val Ile Cys Ser Gly Phe Val Arg Leu Ala Gly Gly Asp Gly Pro Cys Ser Gly Arg Val Glu Val His Ser Gly Glu Ala Trp Thr Pro Val Ser Asp Gly Asn Phe Thr Leu Pro Thr Ala Gln Val Ile Cys Ala Glu Leu Gly Cys Gly Lys Ala Val Ser Val Leu Gly His Met Pro Phe Arg Glu Ser Asp Gly Gln Val Trp Ala Glu Glu Phe Arg Cys Asp Gly Glu Pro Glu Leu Trp Ser Cys

| Pro | Arg 770 | Val | Pro | Cys | Pro | Gly 775 | Gly | Thr | Cys | Leu | His 780 | Ser | Gly | Ala | Ala |
|------------|-------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|-------------|------------|--------------------|------------------------|
| Gln 785 | Val | Val | Cys | Ser | Val 790 | Tyr | Thr | Glu | Val | Gln 795 | Leu | Met | Lys | Asn | G1 ₃ 800 |
| Thr | Ser | Gln | Cys | Glu 805 | Gly | Gln | Val | Glu | Met 810 | Lys | Ile | Ser | Gly | Ar g 815 | Trp |
| Arg | Ala | Leu | Суз 820 | Ala | Ser | His | Trp | Ser 825 | Leu | Ala | Asn | Ala | Asn 830 | Val | Val |
| Суз | Arg | Gln 835 | Leu | Gly | Cys | Gly | Val 840 | Ala | Ile | Ser | Thr | Pro 845 | Arg | Gly | Pro |
| His | Leu 850 | Val | Glu | Gly | Gly | Asp 855 | Gln | Ile | Ser | Thr | Ala 860 | Gln | Phe | His | Cys |
| Ser 865 | Gly | Ala | Glu | Ser | Phe 870 | Leu | Trp | Ser | Cys | Pro 875 | Val | Thr | Ala | Leu | Gl ₃ 880 |
| Gly | Pro | Asp | Cys | Ser 885 | His | Gly | Asn | Thr | Ala 890 | Ser | Val | Ile | Суз | Ser 895 | Gly |
| Asn | His | Thr | Gln 900 | Val | Leu | Pro | Gln | Суз 905 | Asn | Asp | Phe | Leu | Ser 910 | Gln | Pro |
| Ala | Gly | Ser 915 | Ala | Ala | Ser | Glu | Glu 920 | Ser | Ser | Pro | Tyr | Cys 925 | Ser | Asp | Sei |
| Arg | Gln 930 | Leu | Arg | Leu | Val | Asp 935 | Gly | Gly | Gly | Pro | Cys 940 | Gly | Gly | Arg | Va. |
| Glu 945 | Ile | Leu | Asp | Gln | Gly 950 | Ser | Trp | Gly | Thr | Ile 955 | Cys | Asp | Asp | Asp | Trg 960 |
| qeA | Leu | Asp | Asp | Ala 965 | Arg | Val | Val | Суз | Arg 970 | Gln | Leu | Gly | Суз | Gly 975 | G1 |
| Ala | Leu | Asn | Ala 980 | Thr | Gly | Ser | Ala | His 985 | Phe | Gly | Ala | Gly | Ser 990 | Gly | Pro |
| Ile | Trp | Leu 995 | Asp | Asp | Leu | | Cys | Thr | Gly | Lys | | Ser 1005 | His | Val | Tr |
| - | Cys 1010 | Pro | Ser | Arg | _ | Trp | Gly | Arg | His | _ | Cys | Arg | His | Lys | Glu |

Asp Ala Gly Val Ile Cys Ser Glu Phe Leu Ala Leu Arg Met Val Ser 1025 1030 1035 1040

- Glu Asp Gln Gln Cys Ala Gly Trp Leu Glu Val Phe Tyr Asn Gly Thr 1045 1050 1055
- Trp Gly Ser Val Cys Arg Ser Pro Met Glu Asp Ile Thr Val Ser Val 1060 1065 1070
- Ile Cys Arg Gln Leu Gly Cys Gly Asp Ser Gly Ser Leu Asn Thr Ser 1075 1080 1085
- Val Gly Leu Arg Glu Gly Ser Arg Pro Arg Trp Val Asp Leu Ile Gln 1090 1095 1100
- Cys Arg Lys Met Asp Thr Ser Leu Trp Gln Cys Pro Ser Gly Pro Trp 1105 1110 1115 1120
- Lys Tyr Ser Ser Cys Ser Pro Lys Glu Glu Ala Tyr Ile Ser Cys Glu 1125 1130 1135
- Gly Arg Arg Pro Lys Ser Cys Pro Thr Ala Ala Ala Cys Thr Asp Arg 1140 1145 1150
- Glu Lys Leu Arg Leu Arg Gly Gly Asp Ser Glu Cys Ser Gly Arg Val 1155 1160 1165
- Glu Val Trp His Asn Gly Ser Trp Gly Thr Val Cys Asp Asp Ser Trp 1170 1175 1180
- Ser Leu Ala Glu Ala Glu Val Val Cys Gln Gln Leu Gly Cys Gly Gln 1185 1190 1195 1200
- Ala Leu Glu Ala Val Arg Ser Ala Ala Phe Gly Pro Gly Asn Gly Ser 1205 1210 1215
- Ile Trp Leu Asp Glu Val Gln Cys Gly Gly Arg Glu Ser Ser Leu Trp 1220 1225 1230
- Asp Cys Val Ala Glu Pro Trp Gly Gln Ser Asp Cys Lys His Glu Glu 1235 1240 1245
- Asp Ala Gly Val Arg Cys Ser Gly Val Arg Thr Thr Leu Pro Thr Thr 1250 1255 1260
- Thr Ala Gly Thr Arg Thr Thr Ser Asn Ser Leu Pro Gly Ile Phe Ser 1265 1270 1275 1280

Leu Pro Gly Val Leu Cys Leu Ile Leu Gly Ser Leu Leu Phe Leu Val 1285 1290 1295

Leu Val Ile Leu Val Thr Gln Leu Leu Arg Trp Arg Ala Glu Arg Arg 1300 1305 1310

Ala Leu Ser Ser Tyr Glu Asp Ala Leu Ala Glu Ala Val Tyr Glu Glu
. 1315 1320 1325

Leu Asp Tyr Leu Leu Thr Gln Lys Glu Gly Leu Gly Ser Pro Asp Gln 1330 1335 1340

Met Thr Asp Val Pro Asp Glu Asn Tyr Asp Asp Ala Glu Glu Val Pro 1345 1350 1355 1360

Val Pro Gly Thr Pro Ser Pro Ser Gln Gly Asn Glu Glu Glu Val Pro 1365 1370 1375

Pro Glu Lys Glu Asp Gly Val Arg Ser Ser Gln Thr Gly Ser Phe Leu 1380 1385 1390

Asn Phe Ser Arg Glu Ala Ala Asn Pro Gly Glu Gly Glu Glu Ser Phe 1395 1400 1405

Trp Leu Leu Gln Gly Lys Lys Gly Asp Ala Gly Tyr Asp Asp Val Glu 1410 1415 1420

Leu Ser Ala Leu Gly Thr Ser Pro Val Thr Phe Ser 1425 1430 1435

<210> 79

<211> 4308

<212> DNA

<213> Bos sp.

<400> 79

atggetetgg geagacacci etecetgegg ggaetetgtg teeteetet eggeaecatg 60 gtgggtggte aagetetgga getgaggttg aaggatggag teeategetg tgaggggaga 120 gtggaagtga ageaecaagg agaatgggge acagtggatg gttacaggtg gacattgaag 180 gatgeaetetg tagtgtgeag acagetgggg tgtggagetg ecattggtt teetggaggg 240 gettattttg ggeeaggaet tggeeecatt tggettttgt ataetteatg tgaagggaca 300 gagteaactg teagtgaetg tgageattet aatattaaag actategtaa tgatggeta 360 aateatggte gggatgetgg agaagtetge teaggatttg tgegtetgge tggaggggat 420 gggaeectget eagggegagt agaagtgeat tetggagaag ettggateee agtgtetgat 480 gggaaecttea caettgeeae tgeeeagate atetgtgeag agttgggttg tggeaagget 540 gtgetetgee tgggaeagaa geetgagee tgggtetge eecaggtetg ggetgaaga 600 tteaggttg aggggagga geetgagee tgggtetgee eecaggtee etgteeagg 660

| ggcacgtgtc | accacagtgg | atctgctcag | gttgtttgtt | cagcatactc | agaagtccgg | 720 |
|--------------------|--------------------|--------------------|--------------------|------------|--------------------|------|
| ctcatgacaa | acggctcctc | tcagtgtgaa | gggcaggtgg | agatgaacat | ttctggacaa | 780 |
| tggagagcgc | tctgtgcctc | ccactggagt | ctggccaatg | ccaatgttat | ctgtcgtcag | 840 |
| ctcggctgtg | gagttgccat | ctccaccccc | ggaggaccac | acttggtgga | agaaggtgat | 900 |
| cagatectaa | cagcccgatt | tcactgctct | ggggctgagt | ccttcctgtg | gagttgtcct | 960 |
| gtgactgccc | tgggtggtcc | tgactgttcc | catggcaaca | cagcctctgt | gatctgctca | 1020 |
| ggaaaccaga | tccaggtgct | tccccagtgc | aacgactccg | tgtctcaacc | tacaggetet | 1080 |
| gcggcctcag | aggacagcgc | cccctactgc | tcagacagca | ggcagctccg | cctggtggac | 1140 |
| gggggcggtc | cctgcgccgg | gagagtggag | atccttgacc | agggctcctg | gggcaccatc | 1200 |
| tgtgatgacg | gctgggacct | ggacgatgcc | cgcgtggtgt | gcaggcagct | gggctgtgga | 1260 |
| gaagccctca | atgccacggg | gtctgctcac | ttcggggcag | gatcagggcc | catctggttg | 1320 |
| gacaacttga | actgcacagg | aaaggagtcc | cacgtgtgga | ggtgcccttc | ccggggctgg | 1380 |
| gggcagcaca | actgcagaca | caagcaggac | gcgggggtca | tctgctcaga | gttcctggcc | 1440 |
| ctcaggatgg | tgagtgagga | ccagcagtgt | gctgggtggc | tggaagtttt | ctacaatggg | 1500 |
| acctggggca | gtgtctgccg | taaccccatg | gaagacatca | ctgtgtccac | gatctgcaga | 1560 |
| cagcttggct | gtggggacag | tggaaccctc | aactcttctg | ttgctcttag | agaaggtttt | 1620 |
| aggccacagt | gggtggatag | aatccagtgt | cggaaaactg | acacctctct | ctggcagtgt | 1680 |
| ccttctgacc | cttggaatta | caactcatgc | tctccaaagg | aggaagccta | tatctggtgt | 1740 |
| gcagacagca | gacagatccg | cctggtggat | ${\tt ggaggtggtc}$ | gctgctctgg | gagagtggag | 1800 |
| atccttgacc | agggctcctg | ${\tt gggcaccatc}$ | tgtgatgacc | gctgggacct | ggacgatgcc | 1860 |
| cgtgtggtgt | gcaagcagct | gggctgtgga | gaagccctgg | acgccactgt | ${\tt ctcttccttc}$ | 1920 |
| ttcgggacgg | gatcagggcc | catctggctg | gatgaagtga | actgcagagg | agaggagtcc | 1980 |
| caagtatgga | ggtgcccttc | ctggggatgg | cggcaacaca | actgcaatca | tcaagaagat | 2040 |
| gcaggagtca | tctgctcagg | ${\tt atttgtgcgt}$ | ctggctggag | gagatggacc | ctgctcaggg | 2100 |
| cgagtagaag | tgcattctgg | agaagcctgg | accccagtgt | ctgatggaaa | cttcacactc | 2160 |
| cccactgccc | aggtcatctg | tgcagagctg | ${\tt ggatgtggca}$ | aggctgtgtc | tgtcctggga | 2220 |
| cacatgccat | tcagagagtc | ${\tt cgatggccag}$ | gtctgggctg | aagagttcag | gtgtgatggg | 2280 |
| ggggagcctg | agctctggtc | ctgccccaga | gtgccctgtc | caggaggcac | atgtctccac | 2340 |
| agtggagctg | ctcaggttgt | ctgttcagtg | tacacagaag | tccagcttat | gaaaaacggc | 2400 |
| acctctcaat | gtgaggggca | ggtggagatg | aagatctctg | gacgatggag | agcgctctgt | 2460 |
| gcctcccact | ggagtctggc | caatgccaat | gttgtctgtc | gtcagctcgg | ctgtggagtc | 2520 |
| gccatctcca | ccccagagg | ${\tt accacacttg}$ | gtggaaggag | gtgatcagat | ctcaacagcc | 2580 |
| caatttcact | gctcaggggc | tgagtccttc | ctgtggagtt | gtcctgtgac | tgccttgggt | 2640 |
| gggcctgact | gttcccatgg | caacacagec | tctgtgatct | gctcaggaaa | ccacacccag | 2700 |
| gtgctgcccc | agtgcaacga | cttcctgtct | caacctgcag | gctctgcggc | ctcagaggag | 2760 |
| agttctccct | actgctcaga | cagcaggcag | ctccgcctgg | tggacggggg | cggtccctgc | 2820 |
| ggcgggagag | tggagatcct | tgaccagggc | tcctggggca | ccatctgtga | tgatgactgg | 2880 |
| gacctggacg | atgcccgtgt | ggtgtgcagg | cagctgggct | gtggagaagc | cctcaatgcc | 2940 |
| ${\tt acggggtctg}$ | ${\tt ctcacttcgg}$ | ggcaggatca | gggcccatct | ggctggacga | cctgaactgc | 3000 |
| acaggaaagg | agtcccacgt | gtggaggtgc | ccttcccggg | gctgggggg | gcacgactgc | 3060 |
| agacacaagg | aggacgccgg | ggtcatctgc | tcagagttcc | tggccctcag | gatggtgagc | 3120 |
| gaggaccagc | agtgtgctgg | gtggctggag | gttttctaca | acgggacctg | gggcagtgtc | 3180 |
| tgccgcagcc | ccatggaaga | tatcactgtg | tccgtgatct | gcagacagct | tggatgtggg | 3240 |
| gacagtggaa | gtctcaacac | ctctgttggt | ctcagggaag | gttctagacc | ccggtgggta | 3300 |
| gatttaattc | agtgtcggaa | aatggatacc | tctctctggc | agtgtccttc | tggcccatgg | 3360 |
| aaatacagtt | catgctctcc | aaaggaggaa | gcctacatct | catgtgaagg | aagaagaccc | 3420 |
| aagagctgtc | caactgctgc | cgcctgcaca | gacagagaga | agctccgcct | caggggagga | 3480 |
| gacagcgagt | gctcagggcg | ggtggaggtg | tggcacaacg | gctcctgggg | caccgtgtgc | 3540 |
| | | | | | | |

| gatgactcct | ggagcctggc | agaggctgag | gtggtgtgtc | agcagctggg | ctgtggccag | 3600 |
|------------|------------|------------|------------|------------|------------|------|
| | | | ggccctggaa | | | |
| | | | ctgtgggact | | | 3720 |
| cagagcgact | gcaagcacga | ggaggatgct | ggtgtgaggt | gctctggtgt | aaggacaaca | 3780 |
| | | | acctcaaatt | | | |
| ctgcctgggg | ttctctgcct | tatcctgggg | tcgcttctct | tcctggtcct | cgtcatcctg | 3900 |
| gtgactcagc | tactcagatg | gagagcagag | cgcagagcct | tatccagcta | tgaagatgct | 3960 |
| | | | taccttctga | | | |
| agcccagatc | agatgactga | tgtccctgat | gaaaattatg | atgatgctga | agaagtacca | 4080 |
| gtgcctggaa | ctccttctcc | ctctcagggg | aatgaggagg | aagtgccccc | agagaaggag | 4140 |
| gacggggtga | ggtcctctca | gacaggctct | ttcctgaact | tctccagaga | ggcagctaat | 4200 |
| cctggggaag | gagaagagag | cttctggctg | ctccagggga | agaaagggga | tgctgggtat | 4260 |
| gatgatgttg | aactcagtgc | cctgggaaca | tccccagtga | ctttctcg | | 4308 |